تحليل البيانات خطوة بخطوة في SPSS

Data Analysis Step by Step in SPSS

تأليف

د . رامي صلاح جبريل

النسخة الأولى – 2020

IBM® SPSS® Statistics
Version 23

تحليل البيانات خطوة بخطوة في SPSS

Data Analysis Step by Step in SPSS

تأليف

د. رامي صلاح جبريل

أستاذ مشارك في قسم الإحصاء – كلية العلوم

جامعة بنغازي – ليبيا

النسخة الأولى – 2020

النسخة الأولى – 2020

اسم الكتاب: تحليل البيانات خطوة بخطوة في SPSS

اسم المؤلف: د. رامي صلاح جبريل

الوكالة الليبية للترقيم الدولي الموحد للكتاب

دار الكتب الوطنية

بنغازي – ليبيا

هاتف: 9097074 - 9096379 – 9090509

بريد مصور: 9097073

البريد الإلكتروني: nat_lib_libya@hotmail.com

رقم الإيداع القانوني 2019 / 757

ردمك ISBN 978-9959-1-2166-0

تمهيد

بسم الله الرحمن الرحيم

وبه نستعين

الحمد لله رب العالمين، والصلاة والسلام على أشرف المرسلين سيدنا محمد وعلى آله وصحبه أجمعين.

تُعتبر البيانات، من وجهة نظر علم الإحصاء، الشكل المجرد أو الخام الذي تكون عليه المعلومات في الحياة، فهي نتاج قياسات المتغيرات، أي المشاهدات التي تنتج عن تجربة علمية أو تشخيص طبي أو تسجيل لحركة واردات وصادرات في مؤسسة أو نتائج دراسية. وببساطة يمكننا القول إن البيانات هي أي نتيجة يمكن تسجيلها وتحويلها إلى متغير رقمي. وتحليل البيانات، هو ذلك العلم الذي يستخدم الأساليب الإحصائية للتعلم من البيانات، (وهو ما يُعرف حديثا **بالتعلم الإحصائي** (Statistical Learning)). ويُقصد من التعلم هنا استخدام البيانات في الاستكشاف والتحليل للوصول للمعلومات الكامنة، وبالتالي الحصول على المعرفة.

ويتحقق التعلم الإحصائي من خلال تعريف المشكلة أو تحديد الهدف من استخدام تحليل البيانات، ثم الشروع في عملية جمع البيانات، ثم تلخيص واستكشاف وتحليل البيانات باستخدام الأساليب المناسبة، وأخيرا استخلاص النتائج وتفسيرها وعرضها بما يتناسب مع اتخاذ القرار أو تقديم التوصيات والاستشارات المناسبة.

وللوصول للنتيجة الأفضل، يتحتم على الباحث استخدام البرامج الإحصائية التي توفر له ما يساعده في تحقيق أهداف الدراسة، ويُعتبر برنامج IBM SPSS من البرامج المرنة والقوية، وفي نفس الوقت، البسيطة الاستخدام خاصة لغير المتخصصين في علم الإحصاء.

- **لماذا نتعلم برنامج SPSS؟**

برنامج SPSS، والذي يمثل اختصارا لـ (Statistical Package for the Social Sciences)، أي **الحزمة الإحصائية للعلوم الاجتماعية**، يُعد البرنامج الإحصائي المفضل للطلبة والباحثين المتخصصين في العلوم الطبية والنفسية والبيولوجية، والعلوم الاجتماعية والأدبية، وكذلك العلوم الاقتصادية والمالية، وغيرها من العلوم، حيث أنه يحتوي على العديد من الأساليب الإحصائية والمقاييس التي تتعلق بطبيعة هذه التخصصات وطرق تعاملها مع البيانات (المتغيرات) بصورة أحادية (Univariate)، ثنائية (Bivariate)، ومتعددة (Multivariate).

ومن ضمن المزايا التي يتمتع بها برنامج SPSS، مقارنة ببعض البرامج أو الحزم الإحصائية الأخرى الشهيرة مثل برنامج R، (والذي يُعد من وجهة نظر معظم الإحصائيين البرنامج الأفضل حاليا للتحليل الإحصائي)، هي المرونة الفائقة في إدخال البيانات وتعديلها ومعالجتها، وتنوع طرق ترميز وتحوير المتغيرات، وكذلك سهولة استيراد وتصدير ملفات البيانات من وإلى البرامج الأخرى وعلى رأسها برنامج مايكروسوفت اكسل (MS Excel)، والذي يعتبر القاسم المشترك بين كافة برامج معالجة البيانات المتداولة.

وتُعتبر طريقة عرض النتائج في برنامج SPSS من الجوانب المميزة فيه حيث أنها تتميز بالبساطة والاختصار والتدرج المنطقي في العرض إضافة لسهولة تعديلها، (وخاصة نتائج الرسومات البيانية والجداول)، وتعدد طرق نسخها وتخزينها في برامج أخرى مثل مايكروسوفت وورد (MS Word) وباوربوينت (MS Power Point).

- **من يهُم هذا الكتاب؟**

الكتاب يتميز أولا، وكما يظهر من عنوانه، بالتدرج في طرح المواضيع والطرق الإحصائية بصورة بعيدة عن التعقيد والمبالغة في عرض النظريات والقوانين الإحصائية الجافة، فهو يراعي الطلبة والباحثين الذين لديهم بعض الصعوبة في الجانب النظري الرياضي، وكذلك يأخذ في الاعتبار قلة الخبرة الإحصائية لدى البعض، ويتدرج أيضا في شرح كيفية التعامل مع SPSS بطريقة غير معقدة لمن لديهم خبرة متواضعة في استخدام جهاز الحاسوب في العموم. والكتاب لا يشترط وجود أية خبرة سابقة في التعامل مع برنامج SPSS، إلا أنه يفترض أن القارئ لديه إلمام بالحد الأدنى لأساسيات علم الإحصاء الوصفي والاستدلالي، ومفهوم تحليل البيانات. وإضافة لهذا كله، فإن الكتاب يُقدم مجموعة متنوعة من الأمثلة العملية التي تعتمد على بيانات حقيقية وافتراضية تتعلق بالتخصصات العلمية المختلفة؛ الطبية والاجتماعية والتربوية والأكاديمية والاقتصادية والصناعية وغيرها.

- **ما هي التقنيات الإضافية المطلوبة لتعلم برنامج SPSS؟**

للاستفادة من مواضيع الكتاب بصورة جيدة، يُفضل استخدام جهاز حاسوب يعمل تحت نظام تشغيل مايكروسوفت ويندوز (MS Windows) لأي إصدار من ويندوز 7، 8، 8.1 أو 10، ولا تتطلب مواضيع الكتاب وجود اتصال بالإنترنت بعد الحصول على نسخة من برنامج SPSS، والتي يُفضل أن تكون النسخة 23، إلا أن وجود نسخ أقدم لا يمنع القارئ من الاستفادة المثلى من الكتاب.

- **تنظيم فصول الكتاب:**

في **الفصل الأول**، (وهو بعنوان: **مكونات برنامج IBM SPSS**)، نقدم تعريف للبرنامج وطرق استخدام الأوامر فيه، خاصة باستخدام طريقة واجهة المستخدم البيانية (GUI)، ونوضح كيفية طلب المساعدة داخل البرنامج فيما يختص بالمواضيع والتدريب العملي الذي يوفره البرنامج داخل أمر المساعدة. ثم نوضح كيفية إدخال البيانات في ملفات البرنامج مع توضيح مفصل لأسلوبي عرض البيانات وعرض المتغيرات الذي يتميز به برنامج SPSS.

وفي **الفصل الثاني،** (وهو بعنوان: **التعامل مع ملفات البيانات في برنامج SPSS**)، فإننا نبدأ بشرح كيفية استيراد ملفات البيانات من برنامج اكسل وتعديلها ثم تخزينها في SPSS، وبالمقابل نوضح كيفية تصدير أو تخزين الملفات بصيغة اكسل من داخل برنامج SPSS. بعدها يتم شرح كيفية ترتيب المتغيرات في ملفات البيانات في SPSS، وكيفية إعادة ترميز المتغيرات سواء ترميز القيم في نفس المتغير أو في متغير جديد. وكذلك نسلط الضوء على طريقة إعادة الترميز التلقائي وطريقة تعريف وحساب دوال جديدة للمتغيرات.

نبدأ في **الفصل الثالث،** (وهو بعنوان: **التحليل الاستكشافي للبيانات في SPSS**)، بتناول المقاييس والطرق الإحصائية المختلفة التي تندرج تحت مسمى التحليل الاستكشافي للبيانات، ونقوم بتوضيح كيفية استخدامها في برنامج SPSS، ويتضمن الفصل تعريف أنواع البيانات، وتقسيم التعامل مع المتغيرات بحسب طبيعتها، واستخدام مقاييس النزعة المركزية والتشتت والرسومات البيانية بحسب طبيعة المتغيرات. وكذلك توضيح كيفية التعامل مع ما يعرف بالمتغيرات الوهمية، وفي نهاية الفصل الثالث يتعرض الكتاب لاستخدام أسلوب استكشاف البيانات الشامل في SPSS، إضافة لتقديم مثال تطبيقي لكيفية تقديم وعرض نتائج التحليل الاستكشافي للبيانات وتفسيرها.

ونقوم في **الفصل الرابع،** (وهو بعنوان: **تحليل البيانات الاستدلالي في SPSS**)، باستعراض الأساليب الإحصائية التي تتضمن طرق تقدير المعالم واختبارات الفروض المتعلقة بها، مثل الاستدلال حول الوسط الحسابي للمجتمع، والتعامل مع العينات المستقلة والمرتبطة، واختبار تباينات المجتمع، وكذلك نوضح كيفية تكوين جداول الاقتران في SPSS للمتغيرات الوصفية والكمية، وطريقة تنفيذ اختبار الاستقلالية وحساب معاملات الارتباط الخاصة به، وأيضا طرق إجراء اختبارات تحليل التباين في اتجاه واتجاهين.

في **الفصل الخامس،** (وهو بعنوان: **تحليل الارتباط والانحدار الخطي**)، نتناول كيفية تنفيذ تحليل الارتباط والانحدار الخطي باستخدام برنامج SPSS لدراسة العلاقات المتبادلة ضمن المتغيرات في البيانات، وتحليلها، وتكوين النماذج التي تفسر العلاقات السببية بينها. ويتضمن ذلك تنفيذ شكل الانتشار بين المتغيرات في بُعدين أو كمصفوفة، وحساب معاملات الارتباط البسيط والجزئي، ثم شرح طريقة توفيق نموذج الانحدار الخطي البسيط والمتعدد وتمثيلها بيانيا، وبعد ذلك نقدم طرق تقييم نموذج الانحدار الخطي الموفق، وأخيرا نستعرض طرق تقييم أهم فرضيات المربعات الصغرى الاعتيادية وتحليل البواقي.

أما الفصل السادس، (وهو بعنوان: **الاختبارات اللامعلمية وتحليل الموثوقية**)، فيتم فيه التطرق لموضوعين مهمين؛ الأول هو تنفيذ الاختبارات اللامعلمية في SPSS لعينة واحدة، وعينتين أو أكثر (مستقلة أو مرتبطة)، وتتضمن اختبارات هامة مثل اختبار مربع كاي لجودة التوفيق، اختبار ذي الحدين، واختبار الدورات، واختبار كولمجروف–سميرنوف، ومان–وتني، واختبار كروسكال–والس، واختبار ولكوكسون وغيرها. وأما الموضوع الثاني، فيتناول طرق التحقق من ثبات أو موثوقية بيانات الدراسة، (وبصورة خاصة الإجابات في الاستبيانات)، وكيفية تطبيقها وتحليل النتائج في SPSS، ومن أهمها معامل ثبات كرونباخ الفا، ومعامل ثبات التجزئة النصفية، ومعامل ثبات جوتمان.

ويضم الكتاب في نهايته ملحقا يحتوي على تعريف لكل ملفات البيانات المستخدمة بشكل متكرر في الشرح، لكي يتمكن القارئ من مواكبة تنفيذ الأوامر في SPSS خلال فصول الكتاب ولا يضطر للبحث عنها داخل الصفحات.

- **الكتب السابقة للمؤلف:**

1. مقدمة في الإحصاء الاستكشافي والاحتمالات، (2015).
2. التحليل الإحصائي باستخدام لغة R، (2016).

- **شكر واعتزاز:**

أشكر الله العليم الحكيم الذي وفقني بتأليف هذا الكتاب، والذي أرجو أن ينتفع به الطلبة والباحثين في شتى التخصصات، وأحمده عز وجل على هذا العلم، علم الإحصاء، الذي أعتز بأنه الأكثر ذكرا من بين العلوم في القرآن الكريم في أكثر من موضع؛ (... مال هذا الكتاب لا يغادر صغيرة ولا كبيرة إلا **أحصاها** ...، الكهف (49))، (لقد **أحصاهم** وعدهم عدا، مريم (94))، (... **أحصاه** الله ونسوه ...، المجادلة (6))، وغيرها من الآيات الكريمة.

المؤلف

المحتويات

الفصل الخامس 147

تحليل الارتباط والانحدار الخطي
(Linear Correlation and Regression Analysis)

الفصل الأول

مكونات برنامج IBM SPSS
(Components of SPSS)

1.1 تعريف برنامج SPSS (What is SPSS?)

يرمز الاختصار SPSS إلى (Statistical Package for the Social Sciences)، ويعني **الحزمة الإحصائية للعلوم الاجتماعية**، وقد تم ابتكار البرنامج بصورته الأصلية البسيطة في نهاية الستينات عن طريق كل من نورمان ناي (Norman Nie)، هاندلي هل (Hadlai Hull)، ودايل بنت (Dale Bent) من جامعة ستانفورد الأمريكية بغرض تحليل بيانات ذات حجم كبير يصعب التعامل معها بالحسابات اليدوية آنذاك، وبعد ذلك انتشر التعامل بهذه الحزمة في عدة جامعات إلى أن خرج للتداول التجاري لاحقا في الثمانينيات. في عام 2009، استحوذت شركة آي بي إم (IBM) الأمريكية على حزمة SPSS ليصبح بعدها الاسم الرسمي المعروف للبرنامج هو IBM SPSS Statistics.

ويتوفر البرنامج بعدة إصدارات منها إصدار المستخدم المفرد والمتعدد وإصدار الطالب وخادم العميل (clientserver) وغيرها. ويمكن للمستخدم زيارة الموقع الالكتروني الرسمي للشركة، (وهو www.spss.com)، للمزيد من المعلومات.

وهذا الكتاب يستخدم النسخة أو الإصدار IBM SPSS Statistics 23، وإذا ما كان لدى القارئ نسخة أقدم من النسخة 23 أو نسخة أحدث منها مثبتة على جهازه، فإننا نشير هنا إلى أن طريقة استخدام البرنامج وتحليل وعرض النتائج لن يكون فيها اختلاف كبير عن النسخ الأخرى من البرنامج.

ملاحظة:

تجدر الإشارة هنا إلى أن برنامج SPSS كان يُعرف في فترة من الفترات (في الإصدار 18 غالبا)، باسم PASW Statistics وهي اختصار (Predictive Analysis Software) أي **برنامج التحليل التنبئي**.

2.1 طرق استخدام الأوامر (Methods of Commanding)

توجد عدة طرق لاستخدام الأوامر في SPSS بحسب المهمة المطلوب إنجازها أو حتى بحسب ما يفضله المستخدِم، حيث أنه من الممكن استخدام أكثر من طريقة في آن واحد، وهذه الطرق هي:

1. طريقة **واجهة المستخدم البيانية** ((Graphical User Interface (GUI))، وهي واجهة البرنامج المتمثلة في النافذة الأساسية التي يتم عن طريقها استخدام الفأرة واختيار الأوامر من القوائم المنسدلة تحت شريط الأوامر كما هو المتداول في كل البرامج تحت بيئة نظام ويندوز. وهذه الطريقة هي الأكثر تداولا بين مستخدمي البرنامج، وخاصة المبتدئين منهم، نظرا لبساطتها. وتتميز بأنها تعتمد على إدخال الأوامر خطوة بخطوة، بمعنى أن المستخدِم لن يتمكن من الانتقال للخطوة التالية إلا بعد استكمال الإدخال في الخطوة السابقة بصورة صحيحة. وسنقوم في هذا الكتاب باعتماد هذه الطريقة في شرح طريقة التعامل مع البيانات واستخدام التحليل الإحصائي بصورة عامة مع الإشارة للطرق الأخرى عند الضرورة.

2. طريقة **بناء جملة الأوامر** (Syntax)، وهي **لغة الأوامر** (Command Language) الخاصة ببرنامج SPSS، والتي يمكن للمستخدِم من خلالها تنفيذ ما يحتاجه من عمليات حسابية ودوال رياضية وإحصائية عن طريق كتابة أوامر محددة بلغة البرنامج. ويمكن القول بأن استخدام النوافذ والقوائم في الطريقة الأولى، وهي طريقة واجهة المستخدم البيانية، هو في الواقع استخدام للغة الأوامر "في الخلفية". وعادة ما يستخدم هذه الطريقة المستخدمون الأكثر خبرة في برنامج SPSS أو في البرمجة بصورة عامة، وكذلك قد يكون هنالك حاجة لاستخدام هذه الطريقة عند تكوين أوامر أو دوال خاصة غير متوفرة بشكل مباشر في البرنامج.

3. طريقة **بايثون** (Python)، وهي لغة برمجة عامة تحتوي على مكونات خاصة ببرنامج SPSS، ويمكن استخدامها لكتابة برمجيات تعمل ضمن SPSS. ويمكن أيضا استخدام لغة بايثون مع طريقة بناء جملة الأوامر لجعل SPSS ينفذ الدوال والنماذج الإحصائية العامة والخاصة. وتعد هذه طريقة إضافية للاستفادة من برنامج SPSS للباحثين المتخصصين.

4. طريقة **المخطوطات** (Scripts)، وهي برامج مكتوبة بلغة **البيسك** (BASIC) الشهيرة، ويمكن كتابة هذه المخطوطات بطريقة التنفيذ التلقائي (Auto-script) بحيث يتم تنفيذها مباشرة عند الحصول على نتيجة معينة في برنامج SPSS.

1.2.1 الملحقات في برنامج SPSS (Add-on's)

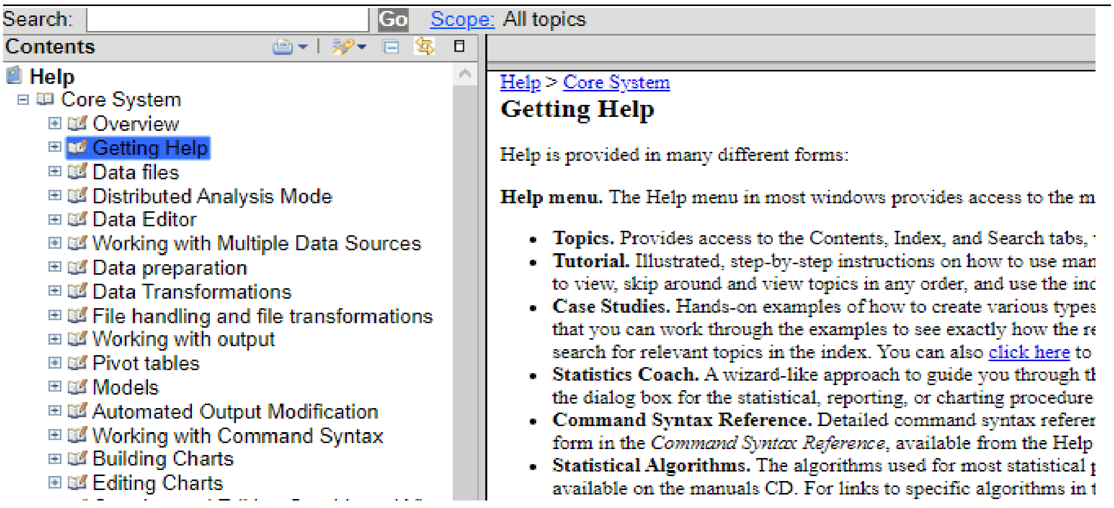
كما هو الحال في كثير من البرامج والحزم الإحصائية تحت بيئة ويندوز، فإنه عند تنصيب برنامج SPSS على الحاسوب يحصل المستخدِم على **النظام الأساسي** (Base System) وربما بعض الملحقات أو الإضافات بحسب طبيعة النسخة التي تم الحصول عليها. ويمكن للمستخدِم الحصول على أية مكونات إضافية

لبرنامج SPSS سواء بتحميلها من الموقع الرسمي للشركة المذكور سابقا، أو من القرص المدمج (CD) الخاص بالبرنامج. وسنقوم بصورة أساسية بتناول الأوامر المتوفرة في النظام الأساسي في SPSS في هذا الكتاب مع التنويه على المكونات الإضافية عند الحاجة لاستخدامها. ويَفترض الكتاب أن القارئ قد حصل مسبقا على نسخة من برنامج SPSS، (سواء النسخة 23 أو أقدم منها أو أحدث)، وأنه قد تم تنصيبها على جهاز الحاسوب، وسيتم خلال فصول الكتاب التدرج في شرح مكونات البرنامج بحسب متطلبات المرحلة.

2.2.1 طلب المساعدة في برنامج SPSS (Getting Help)

يمكن للمستخدم دائما طلب المساعدة في برنامج SPSS في العثور على أي موضوع أو للتعرف على طريقة إدخال معينة للبيانات أو ببساطة لتصفح المواضيع بشكل متسلسل في SPSS، وذلك عن طريق استخدام أمر **المساعدة** (Help) في شريط الأوامر العلوي في البرنامج، وعند استخدام أمر المساعدة، ستتوفر عدة خيارات أهمها ما يلي[1]:

- **المواضيع (Topics)**: ويحتوي على كل الموضوعات التي يوفرها برنامج SPSS متسلسلة بحسب التدرج المنطقي للموضوعات وتسلسل التعامل مع أوامر البرنامج، وتظهر نافذة المساعدة الخاصة بالموضوعات كما يبين الشكل (1.1)، حيث يمكنك التنقل ضمنها مستخدما عناوين المواضيع المعروضة على يسار النافذة، أو عن طريق كتابة الموضوع البحث المطلوب في مربع البحث (Search) في أعلى النافذة إلى اليسار.



شكل 1.1: نافذة طلب المساعدة في المواضيع (Topics) لبرنامج SPSS.

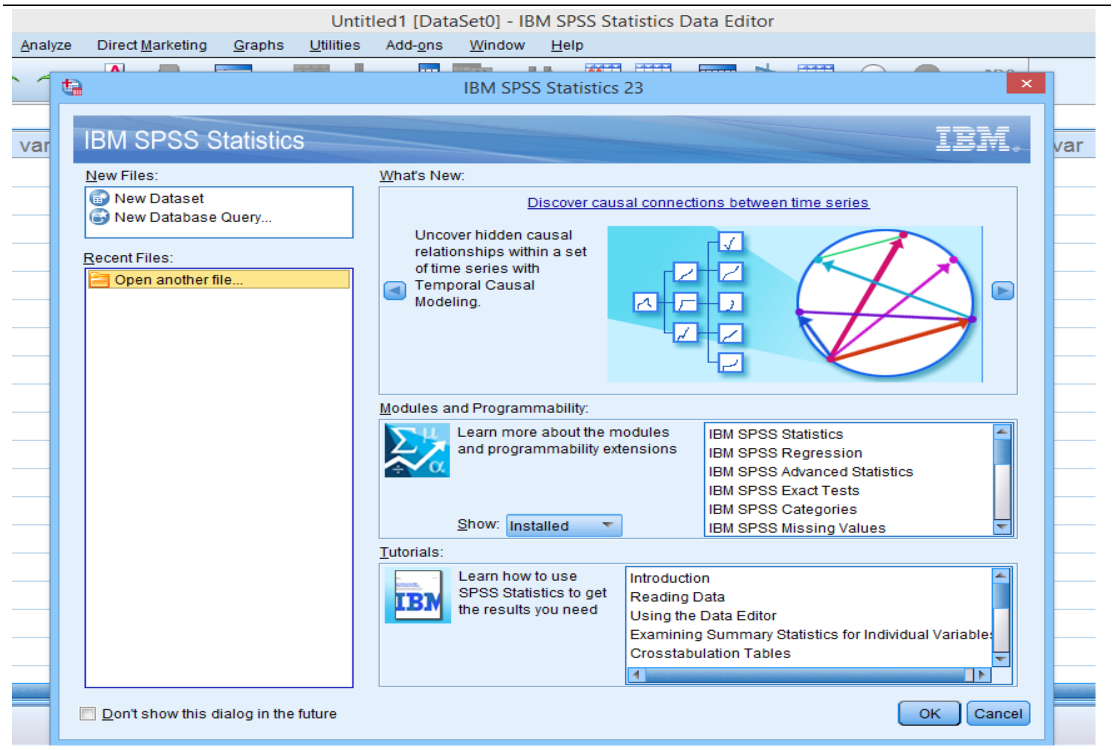
- **التدريب العملي (Tutorial)**: ويوفر للمستخدم نوع من التدريب على استخدام البرنامج في التعامل مع البيانات بصورة عامة لتسهيل تنفيذ التحليل الإحصائي بصورة متدرجة.

[1] عند استخدام أمر المساعدة، فإن عرض المطلوب سيكون عادة في متصفح ملفات الـ pdf أو متصفح الانترنت الافتراضي الخاص بجهازك، لكن بدون الحاجة للاتصال بالإنترنت.

- **دراسات الحالة (Case Studies)**: ويُعد أكثر تقدما من النوعين الأولين في طلب المساعدة، حيث أنه يتضمن أمثلة تطبيقية لبيانات يوفرها البرنامج ويقوم بشرح كيفية استخدام الخيارات المختلفة لبعض الأساليب الإحصائية بصورة عملية.

3.1 بدء التعامل مع برنامج SPSS (Starting SPSS)

عند تشغيل برنامج SPSS للمرة الأولى ستظهر النافذة التالية (شكل (2.1))، والتي تُعتبر النافذة الافتراضية التي يقوم البرنامج بعرضها مقدما فيها بعض الاقتراحات للمستخدم؛ مثل فتح ملف بيانات جديد أو سابق، عرض بعض التحليلات الإحصائية المقترحة، عرض المكونات الموجودة وغير الموجودة في البرنامج، عرض ملاحظات إرشادية وتعليمية لاستخدام البرنامج، وغيرها. وستجد في الزاوية اليسرى أسفل هذه النافذة خيار خاص بتعطيل ظهور هذه النافذة عند فتح البرنامج مرة أخرى، وسنقوم باختيار هذا التعطيل لعدم حاجتنا لهذه النافذة خلال الفصول الأولى، وذلك عن طريق الضغط على المربع الصغير بجانب الخيار ثم الضغط على زر موافق (OK) في النافذة.



شكل 2.1: نافذة "الترحيب" الافتراضية عند تشغيل برنامج **SPSS**.

بعد تنفيذ ذلك، ستبقى النافذة الرئيسية للبرنامج مفتوحة، والشكل (3.1) يوضح أهم المكونات الأساسية لهذه النافذة التي تُعد واجهة المستخدم البيانية (GUI) التي تكلمنا عنها سابقا في طرق استخدام الأوامر في البرنامج، وتسمى نافذة **تحرير البيانات** (SPSS Data Editor)، وستكون بالنسبة لنا نافذة التعامل مع البرنامج عن طريق إدخال البيانات وتعديلها، وإصدار الأوامر الخاصة بالتحليل الإحصائي من مقاييس ونماذج ورسومات بيانية وغيرها.



شكل 3.1: نافذة تحرير البيانات في برنامج SPSS.

في الشكل (3.1) نلاحظ أن نافذة تحرير البيانات مقسمة إلى:

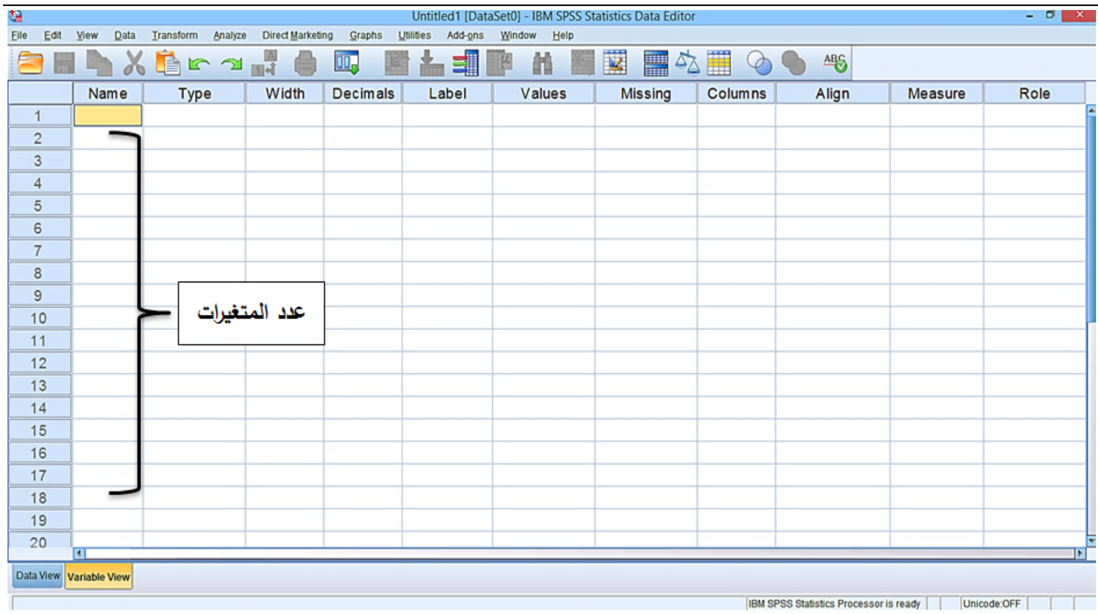
- **اسم ملف البيانات:** حيث يظهر (إلى اليسار) اسم ملف البيانات (Untitled1) وهو الاسم الافتراضي الذي يعطيه البرنامج لملفات البيانات، ويقوم المستخدِم بتغييره لاحقا إلى الاسم الذي يرغب به بعد إدخال البيانات.
- **شريط الأوامر:** وهو الشكل العام لعرض الأوامر الأساسية في بيئة ويندوز، وعند الضغط على أي من هذه الأوامر ستنسدل قائمة تضم أوامر فرعية أو تضم قوائم تحوي بدورها أوامر فرعية. وسنقوم بالتعرف على هذه الأوامر بالتدريج بحسب تقدم القارئ مع فصول الكتاب.
- **أيقونات الأوامر:** وهي رموز تنفيذية للأوامر وتعتبر بدائل أو اختصارات لها، ويمكن تعديل هذا الشريط ليضم الأيقونات المفضلة أو الأكثر استخداما بحسب رغبة المستخدِم.
- **المتغيرات والمشاهدات:** كما هو السائد في التحليل الإحصائي، فإن الشكل العام لملف البيانات الإحصائية يتم تنظيمه بحيث تكون المتغيرات في الأعمدة والمشاهدات في الصفوف[1]. ويُلاحظ في نافذة تحرير البيانات أن المتغيرات كلها تأخذ الاسم المختصر (var) نظرا لأنه لم يتم تعريفها بعد.

ويجب التنويه هنا إلى أسلوبي العرض الرئيسيين في نافذة تحرير البيانات والمتوفرتان في أسفل جدول البيانات مباشرة إلى اليسار؛ وهما أسلوب **عرض البيانات** (Data View)، وهي الظاهرة في الشكل السابق، وأسلوب **عرض المتغيرات** (Variable View)، والتي عند اختيارها سيتغير الشريط الذي يضم أسماء المتغيرات إلى تقسيم آخر خاص بتصنيف المتغيرات بحسب طبيعتها، وسنتناول تاليا هذين الأسلوبين بالشرح.

[1] سيتم تناول مفهوم المتغيرات والمشاهدات بمزيد من التفصيل في الفصل التالي.

في نافذة محرر البيانات، قم باختيار اسلوب عرض المتغيرات (Variable View) في أسفل النافذة[1] فيتغير العرض في النافذة إلى الأسلوب الموضح في الشكل (4.1). عند إدخال البيانات لأول مرة في محرر البيانات، يتم تعريف أسماء المتغيرات المطلوب إدخالها ونوعها والسمات التفصيلية الخاصة بها باستخدام اسلوب عرض المتغيرات، بعد ذلك يتم الرجوع إلى اسلوب عرض البيانات لإدخال القيم العددية لهذه المتغيرات.

ولاحظ أن الصفوف في نافذة عرض المتغيرات تمثل في الواقع عدد المتغيرات، (وليس كما هو الحال في نافذة عرض البيانات، حيث أنها تمثل عدد المشاهدات).



شكل 4.1: اسلوب عرض المتغيرات في نافذة محرر البيانات في SPSS.

وفيما يلي، نقدم وصف موجز لما تعنيه المصطلحات الخاصة بأسلوب عرض المتغيرات:

- **اسم المتغير** (Name): وهو الاسم الذي يختاره المستخدِم للمتغير، ويمكن استخدام اللغة العربية أو الإنجليزية[2] لإدخال أسماء المتغيرات، مع ملاحظة أنه لا يمكن استخدام المسافات بين الأحرف أو استخدام بعض الرموز ضمن اسم المتغير مثل (&) أو ابتداء اسم المتغير برقم[3]. وفي حال عدم كتابة اسم محدد للمتغير من قِبل المستخدِم، فإنه يتم إعطاء الاسم الافتراضي "VAR00001" للمتغير الأول، والاسم "VAR00002" للمتغير الثاني، ....، وهكذا.

[1] في إصدار SPSS 23 ستضيئ خلفية أسلوب العرض المختار (وكذلك عند الضغط بالفأرة على أي أمر أو خيار) باللون الأصفر لتنبيه المستخدِم.

[2] يمكن التعرف على اللغات التي يمكن استخدامها في الإدخال في برنامج SPSS من الموقع الالكتروني الرسمي للشركة.

[3] حاول مثلا ادخال أسماء المتغيرات التالية؛ متغير "مسافة 1"، أو 5 "مسافة" X، أو 3Y، أو A&B، وستلاحظ ظهور رسالة تنبيه من البرنامج بأن هذا الاسم غير مقبول للمتغير.

- **سمة أو طبيعة المتغير** (Type): والذي يمكنك من اختيار سمة المتغير، (الذي سيأخذ على أية حال الشكل الرقمي وليس النصي)، من ضمن عدة سمات مثل؛ **رقمي** (Numeric) وهي السمة الافتراضية، **مصحوب بفواصل** (Comma) مثل "13,150,719"، **مصحوب بنقاط** (Dot) مثل "13.150.719"، **علمي** (Scientific notation) مثل "1.23E+2" والذي يعني "1.23×10^2"، وغيرها من السمات التي يمكن للمستخدِم الاختيار منها.
- **عدد الخانات العشرية** (Decimals): ويتم فيها تحديد عدد الخانات العشرية المرغوبة للأعداد، وتكون عادة 0 للأعداد الصحيحة، والعدد الافتراضي للخانات العشرية هو 2، والذي يمكنك تغيره من نافذة الإعدادات[1] العامة في البرنامج.
- **مسافة (عرض) قيم المتغير** (Width): ويتم فيه تحديد قياس العرض المطلوب كمسافة لقيم المتغير مع أخذ عدد الخانات العشرية بالاعتبار، فمثلا إذا كان المطلوب كتابة خانتين عشريتين إلى جانب العدد الصحيح فيجب ألا يقل قياس عرض قيم المتغير عن 3. ويكون القياس الافتراضي لعرض قيم المتغير هو 8.
- **طابع أو وصف المتغير** (Label): والذي يمكنك من إعطاء وصف اسمي للمتغير للمساعدة في فهم ما يمثله هذا المتغير. ويمكن أن يكون الوصف باللغة العربية، وأيضا يمكن هنا استخدام المسافات والرموز بمعنى أنه يمكن للمستخدِم كتابة ما يريده بدون قيود. وهذا الوصف هو ما يتم عرضه كبديل عن اسم المتغير في نتائج التحليل، وفي حالة عدم إضافة تعريف للمتغير فإن هذه الخانة تبقى فارغة افتراضيا.
- **تصنيف قيم المتغير** (Values): في هذه الخانة، يتم تعريف أو إعطاء وصف لمستويات أو تقسيمات المتغير الوصفي[2]، فمثلا إذا كان المتغير يمثل إجابة على سؤال معين فإنه يمكن تعريف "موافق" لقيمة المتغير "1"، و"غير موافق" لقيمة المتغير "2". ويمكن هنا أيضا استخدام المسافات والرموز في التعريف. أما في حال كون المتغير كميا أو عند عدم تعريف أي تصنيف للمتغير الوصفي، فإن الخانة يظهر فيها الوصف **لا يوجد** (None).
- **تعريف القيم المفقودة** (Missing): في بعض الأحيان قد يرغب المستخدِم بالتمييز بين قيم مفقودة فعليا وأخرى ناتجة عن عدم الرغبة في الإجابة، كما هو الحال مثلا في بعض الاستبيانات؛ فأحيانا قد يتم فقد بعض استمارات الاستبيان وهذا يندرج تحت مسمى القيم المفقودة فعليا، وأحيانا قد لا يرغب الشخص المستبين في الإجابة عن سؤال معين فتنتج قيمة مفقودة بسبب عدم الاستجابة. في مثل هذه الحالات، يمكن للمستخدِم استخدام هذه الخانة لتعريف القيم المفقودة كما يرغب. وفي حالة عدم وجود قيم مفقودة في المتغير فإن الخانة يظهر فيها الوصف لا يوجد (None).

---

[1] يمكنك التعرف على نافذة الإعدادات العامة في البند (5.1).

[2] سيتم التطرق لمفهوم المتغيرات الكمية والوصفية بالتفصيل في الفصل القادم.

- **عرض العامود** (Columns): ويمثل المساحة التي ستشغلها قيم المتغير ضمن العامود، وهو خيار شكلي فقط لا يؤثر تغييره في تغيير قيم المتغير. ويكون عرض العامود الافتراضي للمتغير هو 8.
- **محاذاة أو تراصف قيم المتغير** (Align): وهو الخيار الخاص بمحاذاة قيم المتغير إلى اليمين (Right) ضمن العامود أو اليسار (Left) أو المنتصف (Center). علما بأن الوضع الافتراضي للمحاذاة يكون إلى اليمين.
- **نوع أو قياس المتغير** (Measure): في هذه الخانة يتم تحديد نوع المتغير من بين ثلاثة خيارات؛ **كمي** (Scale) أو **وصفي رتبي** (Ordinal) أو **وصفي اسمي** (Nominal). أما إذا لم يتم تحديد نوع المتغير من قِبل المستخدِم فسيظهر الوصف (Unknown) بمعنى أن نوع المتغير غير محدد.
- **وظيفة أو دور المتغير** (Role): في معظم الأساليب الإحصائية، يكون للمتغير أو المتغيرات دور معين في الأسلوب أو النموذج الإحصائي[1]، فيمكن أن يكون دور المتغير مُدخل (Input) أو هدف (Target) أو يأخذ الدورين بحسب طبيعة الأسلوب الإحصائي (Both) أو ليس له دور محدد مسبقا (None) أو له دور تقسيمي أو تصنيفي[2] (Partition) أو دور لتجزئة البيانات (Split). وتعيين هذا الدور يكون عادة لدواعي تنظيمية فقط ولا يؤثر ترك المتغيرات بدون تعيين على دورها في سير مرحلة التحليل الإحصائي في العموم. والتعيين الافتراضي لدور المتغير يكون كمُدخل.

4.1 إدخال البيانات في نافذة تحرير البيانات (Data Entry in SPSS Data Editor)

في هذا البند، سيتم التعرف على كيفية إدخال البيانات (أو تعريف المتغيرات) في نافذة تحرير البيانات من خلال تناول مثال افتراضي بسيط يشمل إدخال بيانات جديدة بصورة مباشرة، علما بأنه توجد عدة طرق لإدخال البيانات أو استدعائها إلى برنامج SPSS، والتي سيتم التطرق إليها لاحقا.

عند استخدام الطريقة التقليدية لواجهة المستخدم البيانية، وهي كما أسلفنا الطريقة الأكثر شيوعا بين المستخدمين الجدد في إدخال البيانات، فإن الإدخال يكون عادة على مرحلتين؛ المرحلة الأولى هي تعريف أسماء المتغيرات ونوعها وخواصها (ويتم ذلك باستخدام نافذة أسلوب عرض المتغيرات (Variable View))، والمرحلة الثانية هي التي يتم فيها إدخال قيم هذه المتغيرات التي تم تعريفها، (باستخدام نافذة أسلوب عرض البيانات (Data View)). وسنقوم تاليا بالتدرب على تنفيذ هاتين المرحلتين.

[1] في أسلوب تحليل الانحدار على سبيل المثال، (والذي سيتم تناوله في الفصل الخامس)، يتم تعريف دور أحد المتغيرات بأنه تابع وتعريف دور المتغير أو المتغيرات الأخرى بأنها توضيحية.

[2] بمعنى أن يتم تقسيم بيانات الدراسة إلى ثلاثة أقسام؛ عينة للتحليل (Training Sample)، وعينة اختبارية (Testing Sample)، وعينة للتقييم (Validation) كما هو الحال في بعض الدراسات الإحصائية.

1.4.1 نافذة عرض المتغيرات (Variable View Window)

البيانات في الجدول (1.1) هي بيانات افتراضية خاصة بعشرة أطفال في إحدى المؤسسات التعليمية في مرحلتي رياض الأطفال والابتدائية، وتشمل ثلاثة متغيرات هي وزن الطفل وعمره وجنسه. وسنقوم بإدخال هذه البيانات في ملف بيانات في برنامج SPSS بالطريقة التقليدية المباشرة.

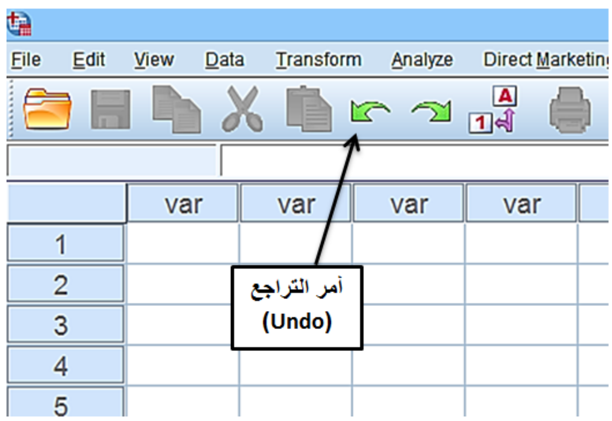
جدول 1.1: بيانات افتراضية خاصة بعشرة أطفال في إحدى المؤسسات التعليمية.

وزن الطفل (بالكجم)	10.2	12	9.5	15.2	20.1	18.9	16.4	13	21.6	14.8
عمر الطفل (بالسنة)	4	5	4	6	8	7	6	5	9	6
جنس الطفل	ذكر	ذكر	أنثى	ذكر	أنثى	أنثى	ذكر	أنثى	أنثى	ذكر

في المرحلة الأولى، وبعد تشغيل البرنامج، تأكد أولا من ظهور نافذة المستخدِم البيانية "الفارغة" كما هو موضح في الشكل (2.1) السابق. قم بعد ذلك باختيار أسلوب عرض المتغيرات (Variable View) من أسفل النافذة إلى اليسار فتظهر النافذة كما في الشكل (4.1) السابق. في هذه النافذة، قم بالضغط باستخدام الفأرة على أول مربع فارغ تحت اسم المتغير (Name) وستلاحظ تغير لون الخلفية إلى اللون الأصفر.

ملاحظة:

في حال حدوث أي خطأ أثناء عملية الإدخال، يمكن للمستخدِم، كما هو الحال في معظم البرامج تحت بيئة ويندوز، أن يستخدم أيقونة (أمر) **التراجع** (Undo) المشار إليها في الشكل (5.1) للرجوع عن هذا الخطأ ثم إعادة الإدخال بالصورة الصحيحة.



شكل 5.1: أمر التراجع (Undo) في نافذة محرر البيانات في SPSS.

سنقوم الآن بإدخال اسم المتغير الأول في هذا الموضع، وكما تمت الإشارة سابقا، يمكنك اختيار الاسم المرغوب ضمن القيود المحددة. ليكن اسم المتغير الأول هو "الوزن"، قم بكتابة هذا الاسم في المربع الأول، ثم اضغط على مفتاح الإدخال[1] في لوحة مفاتيح حاسوبك (ENTER). ستلاحظ هنا أن باقي الخانات في نفس الصف الأول قد ظهرت فيها قيم أو مصطلحات، وهي كما ذكرنا القيم الافتراضية المصاحبة لإدخال أي متغير جديد.

[1] يمكنك أيضا بدلا من استخدام مفتاح الإدخال الضغط بالفأرة على أي مكان فارغ في النافذة لتنفيذ الإدخال.

وهنا يكون للمستخدِم الخيار إما بتعديل بعض أو حتى كل القيم والخيارات للمتغير الجديد بحسب ما يناسب هذا المتغير، أو أن يستمر بكتابة أسماء المتغيرات الجديدة الأخرى في عامود الأسماء ومن ثمة يعود لتعديل الخيارات للمتغيرات المُدخلة. سنقوم نحن باعتماد الخيار الأول، أي أنه سيتم تعديل الخيارات بحسب أولوية إدخال المتغيرات. في الصف الأول، سنقوم بتعديل الخيارات التالية:

- الخانات العشرية (Decimals): قم بكتابة[1] الرقم 1 ثم اضغط إدخال، أي اختيار خانة عشرية واحدة حيث أن ذلك يتناسب مع طبيعة القيم العشرية لهذا المتغير.
- وصف المتغير (Label): يمكن كتابة[2] الوصف التالي؛ "وزن الطفل بالكجم" ثم اضغط إدخال[3].
- نوع المتغير (Measure): قم باختيار نوع المتغير الكمي (Scale) والذي يتناسب مع الطبيعة الكمية لمتغير الوزن.

نلفت انتباه القارئ هنا إلى أن باقي الخيارات الخاصة بالمتغير قد تم تجاهلها إما لعدم الحاجة إليها (مثل تصنيف قيم المتغير (Values) والقيم المفقودة (Missing))، أو لكونها خيارات ثانوية يمكن تعديلها أو تركها على حالتها الافتراضية (مثل ما تبقى من الخيارات الأخرى).

ننتقل الآن لتعريف المتغير الثاني، وهو عمر الطفل؛ قم بكتابة اسم المتغير "العمر" في الخانة الثانية في العامود الأول إلى اليسار، تحت اسم المتغير الأول "الوزن" مباشرة، وبالمثل ستجد أن خانات الخيارات في الصف الثاني قد ظهرت بقيمها الافتراضية، وسنقوم بتعديل الخيارات التالية:

- الخانات العشرية (Decimals): قم بكتابة الرقم 0 ثم اضغط إدخال، حيث أن هذا المتغير لا يتطلب أية خانات عشرية.
- وصف المتغير (Label): قم بكتابة الوصف التالي؛ "عمر الطفل بالسنوات" ثم إدخال.
- نوع المتغير (Measure): قم باختيار نوع المتغير الكمي (Scale) والذي يتناسب مع الطبيعة الكمية لمتغير العمر.

وأخيرا، في أسلوب عرض المتغيرات، نقوم بتعريف المتغير الثالث في البيانات، وهو جنس أو نوع الطالب بنفس الطريقة؛ قم بكتابة اسم المتغير "النوع" في الخانة الثالثة في العامود الأول إلى اليسار، تحت اسم المتغير الثاني،

[1] يمكنك أيضا استخدام السهمين اللذين سيظهران داخل الخانة عند الضغط مرتين بالفأرة لزيادة أو إنقاص العدد المطلوب، وهذين السهمين سيظهران في كل الخانات التي تضم قيم عددية في العموم في أسلوب عرض المتغيرات (Variable View).

[2] في حال توفر وصف جاهز (مكتوب مسبقا) للمتغير في ملف وورد (Word) أو أي ملف بصيغة "txt" يمكن أخذ نسخة منه (Copy) ولصقها (Paste) داخل الخانة المطلوبة.

[3] في هذه المرحلة من الكتاب نذكّر القارئ دائما بالضغط على مفتاح الإدخال بعد كل عملية كتابة للقيم والخيارات، إلا أننا سنترك هذا التذكير في المراحل القادمة بعد أن يكون القارئ قد اعتاد على عملية الإدخال.

وبالمثل ستجد أن خانات الخيارات في الصف الثالث قد ظهرت بقيمها الافتراضية، وستكون تعديلات الخيارات بالصورة التالية:

- الخانات العشرية (Decimals): قم بكتابة الرقم 0 ثم اضغط إدخال، حيث أن هذا المتغير لا يتطلب أية خانات عشرية.
- وصف المتغير (Label): قم بكتابة الوصف التالي؛ "نوع أو جنس الطفل" ثم إدخال.
- نوع المتغير (Measure): قم باختيار نوع المتغير الوصفي الاسمي (Nominal) والذي يتناسب مع طبيعة[1] متغير النوع.
- تصنيف قيم المتغير (Values): قم بالضغط مرتين على هذه الخانة فيظهر مربع صغير باللون الأزرق في يمين الخانة، قم بالضغط عليه فتظهر نافذة صغيرة في وسط الشاشة كما هو موضح في الشكل (6.1). سنقوم الآن بإعطاء القيمة 1 كرمز للذكور، والقيمة 2 كرمز للإناث، فنبدأ بكتابة القيمة 1 في الخانة المقابلة للقيمة (Value)، وكتابة "ذكر" في خانة الوصف (Label) أسفل منها بدون استخدام أقواس في كتابة الوصف.

بمجرد الانتهاء من تعبئة هاتين الخانتين، سيضاء المربع الخاص بالأمر (Add) أسفل منهما. قم بالضغط عليه فيظهر في الخانة المقابلة له الوصف (1="ذكر")، بمعنى أنه قد تم تعريف القيمة 1 بأنها تمثل أو تصف التصنيف "ذكر" في المتغير المسمى النوع[2].

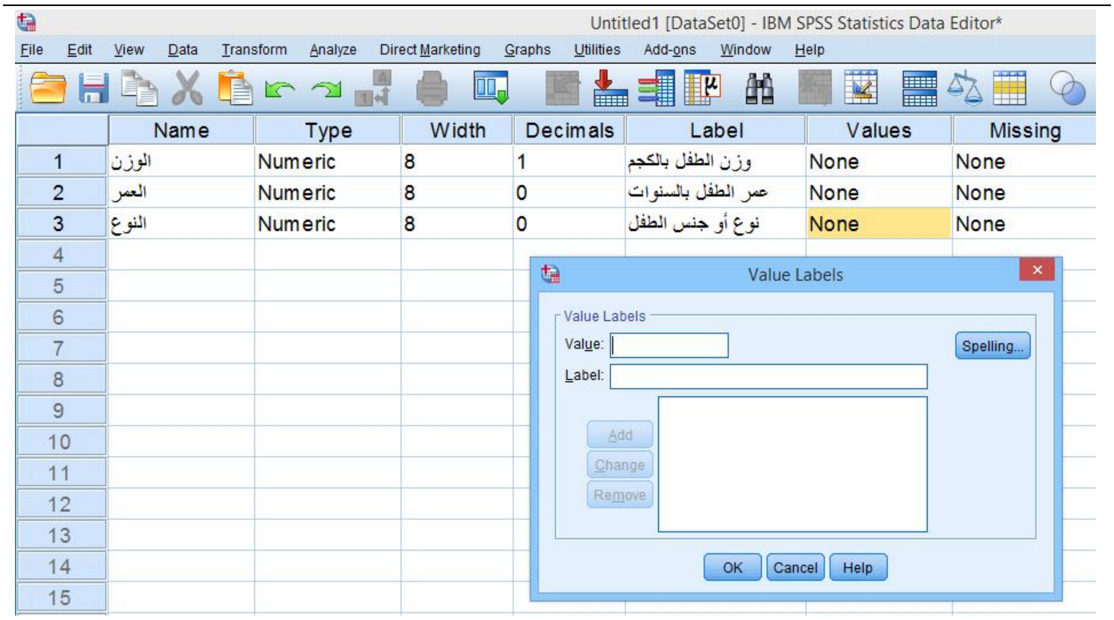
بالمثل سنقوم بتعريف الوصف الثاني، وهو الإناث بنفس الطريقة، قم بكتابة القيمة 2 في الخانة المقابلة للقيمة (Value)، وكتابة "أنثى" في خانة الوصف (Label) ثم الضغط على الأمر (Add) فيظهر في الخانة المقابلة له الوصف (2="أنثى") أسفل الوصف الأول[3]. وبعد الانتهاء من تصنيف قيم المتغير، نضغط أمر موافق (OK) في نفس النافذة لتأكيد الانتهاء من التصنيف.

---

[1] سيتم توضيح كيفية التمييز بين أنواع المتغيرات الكمية منها والوصفية في الفصل التالي.

[2] في حال حدوث أي خطأ أثناء تعريف القيم للمتغير، يمكنك استخدام المربع الخاص بالأمر (Change) للإجراء التعديل المطلوب، أو استخدام الأمر (Remove) لحذف الوصف الموجود وإعادة كتابة وصف جديد.

[3] في حال وجود أكثر من تصنيف للمتغير الوصفي، يتم الاستمرار بإدخال القيم العددية بنفس التسلسل حتى استكمال كل التصنيفات الموجودة، فمثلا إذا كان المتغير الوصفي يمثل تقديرات طلبة هي؛ (راسب، مقبول، جيد، جيد جدا، ممتاز)، فإنه يتم تعريف القيم التالية؛ (1، 2، 3، 4، 5) مقابل كل تصنيف.

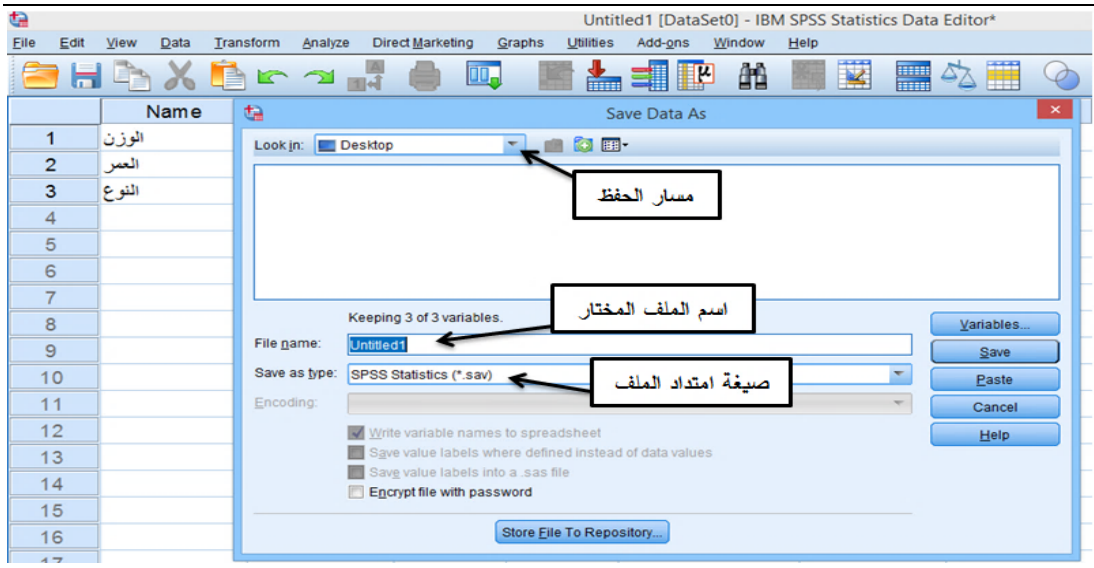


شكل 6.1: النافذة الخاصة بتعريف قيم المتغير الوصفي في نافذة محرر البيانات في برنامج SPSS.

وبهذا نكون قد انتهينا من تعريف المتغيرات الثلاثة، ويمكن البدء بإدخال قيم هذه المتغيرات في أسلوب عرض البيانات (Data View)، إلا أنه من الأفضل دائما حفظ ما تم إنجازه حتى الآن قبل المضي قدما في إدخال قيم المتغيرات، تجنبا لفقدان البيانات نتيجة حدوث خلل طارئ في جهاز الحاسوب أو غيرها من الأسباب.

2.4.1 حفظ ملفات البيانات (Saving Data Files)

في شريط الأوامر العلوي، (المشار إليه في شكل (3.1) سابقا)، قم باختيار أمر **ملف** (File)، فتظهر قائمة أسفل منه تحتوي على بعض الأوامر الفرعية. قم باختيار أمر **الحفظ** (Save)، فتظهر نافذة الحفظ كما هو موضح في الشكل (7.1).

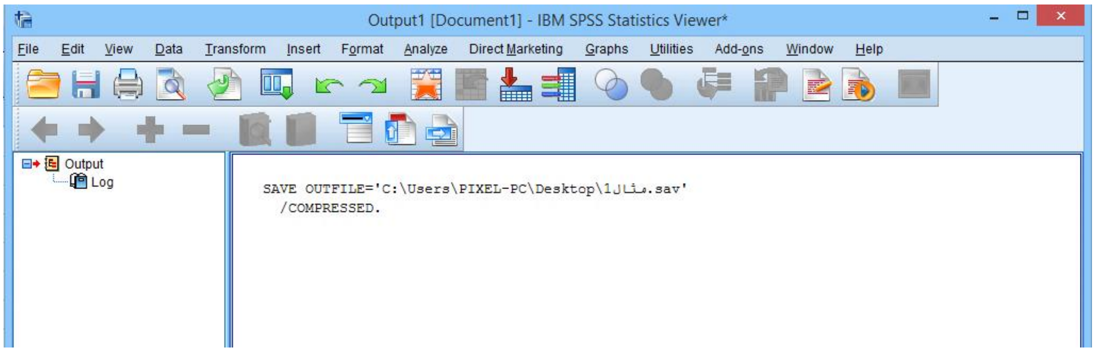


شكل 7.1: نافذة حفظ الملفات في برنامج SPSS.

في شريط **مسار الحفظ** (Look in) في أعلى نافذة الحفظ، قم باختيار المسار أو الموقع الذي ترغب في حفظ ملف البيانات فيه، ولتسهيل حفظ واستدعاء الملفات الخاصة ببرنامج SPSS نقترح على القارئ استخدام مسار **سطح المكتب** (Desktop) حيث أنه المسار الأكثر استخداما تحت بيئة ويندوز في العموم.

في شريط اختيار **اسم الملف** (File name)، يمكنك اختيار الاسم الذي ترغب به لملف البيانات وسنقوم نحن باختيار الاسم "مثال1". ويقوم برنامج SPSS بتعيين الامتداد ".sav" لملفات البيانات فيه، كما يقوم بتعيين امتدادات أخرى للأنواع الأخرى من الملفات كما سنرى لاحقا.

بعد الانتهاء من تحديد مسار الحفظ وكتابة اسم ملف البيانات، قم بالضغط على أمر الحفظ (Save) في يمين نافذة الحفظ فتظهر لك نافذة جديدة كما في الشكل (8.1).



شكل 8.1: نافذة المخرجات في برنامج SPSS.

هذه النافذة تعرف **بنافذة المخرجات** (Output) في برنامج SPSS، وهي تظهر بشكل منفصل بمعنى أن المستخدِم سيتعامل مع هذه النافذة في إطار جديد منفصل عن إطار البرنامج الرئيسي كما يُشاهد في شريط المهام (Taskbar) أسفل الشاشة. وهذا هو الأسلوب القياسي المتبع في برنامج SPSS في عرض كل النتائج والمخرجات الناتجة عن الأوامر المدخلة. وهذه النافذة ستأخذ الاسم "Output1"، وهو الاسم الافتراضي لملف المخرجات، ويمكن للمستخدِم حفظ ملف المخرجات بصورة مستقلة كما هو الحال مع ملف البيانات.

الآن سيُلاحظ القارئ ظهور رسالة داخل نافذة المخرجات توضح ما تم من حفظ لملف البيانات في مسار سطح المكتب باسم " مثال1". في الوقت الحالي، قم بتصغير نافذة المخرجات للاحتفاظ بها أسفل الشاشة، والعودة للنافذة الرئيسية للبرنامج حيث ستلاحظ أن اسم ملف البيانات الجديد قد ظهر في الشريط العلوي للبرنامج.

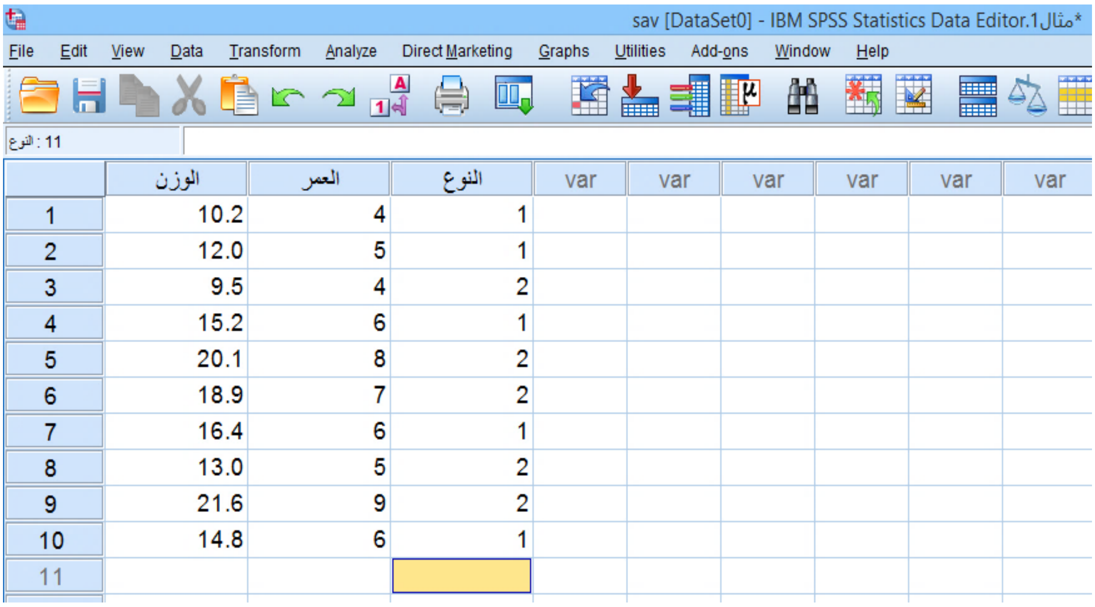
3.4.1 نافذة عرض البيانات (Data View Window)

قبل البدء بعملية الحفظ، كنا قد انتهينا من تعريف المتغيرات للمثال الافتراضي في جدول (1.1) السابق، والآن ننتقل للمرحلة الثانية وهي إدخال قيم هذه المتغيرات في SPSS. قم من أسفل النافذة الرئيسية باختيار نافذة

اسلوب عرض البيانات (Data View)، وستلاحظ أن المتغيرات الثلاثة الأولى إلى اليسار قد أخذت أسماؤها كما تم إدراجها في أسلوب عرض المتغيرات؛ الوزن، العمر ، والنوع.

قم بإدخال قيم المتغير الأول، وهو الوزن بنفس ترتيبها في الجدول (1.1)، أي ستبدأ بإدخال القيمة 10.2 ثم 12 وهكذا وصولا للقيمة الأخيرة 14.8. وفي حل إدخال قيمة بصورة خاطئة في إحدى الخلايا (المربعات الفارغة) يمكنك العودة لتلك الخلية وإعادة كتابة القيمة الصحيحة.

بعد ذلك يتم إدخال قيم المتغير الثاني وهو العمر بنفس الكيفية، أي بداية من القيمة 4 ثم 5 وصولا للقيمة الأخيرة في المتغير وهي 6. أما بالنسبة للمتغير الوصفي النوع، فإن قيمه يتم إدخالها بصورة رقمية وليس اسمية كما تم تعريفها في خانة الوصف (Label) في أسلوب عرض المتغيرات في المرحلة الأولى، أي أننا سندخل القيمة 1 مقابل كل نوع "ذكر" والقيمة 2 مقابل كل نوع "أنثى". وفي النهاية ستكون القيم في ملف البيانات لديك كما هو موضح في الشكل (9.1).



	الوزن	العمر	النوع	var	var	var	var	var	var
1	10.2	4	1						
2	12.0	5	1						
3	9.5	4	2						
4	15.2	6	1						
5	20.1	8	2						
6	18.9	7	2						
7	16.4	6	1						
8	13.0	5	2						
9	21.6	9	2						
10	14.8	6	1						
11									

شكل 9.1: قيم المتغيرات في ملف البيانات "مثال1".

ويمكن للمستخدِم الضغط على أيقونة عرض وصف المتغير  في شريط أيقونات الأوامر للتغيير بين عرض قيم المتغير الرقمية أو قيم المتغير الاسمية.

ملاحظة:

لاحظ ظهور رمز النجمة (*) بجانب اسم ملف البيانات في الشريط العلوي، وهذه هي طريقة البرنامج في تنبيه المستخدِم إلى أنه قد تم إجراء تعديلات أو إضافات إلى ملف البيانات ولم يتم حفظها بعد، لذلك يمكنك بعد إتمام إدخال البيانات حفظ الملف باستخدام File>Save بمعنى الضغط على ملف (File) في شريط الأوامر ثم حفظ (Save)، أو لمزيد من السرعة، يمكنك الضغط على أيقونة الحفظ فيتم حفظ التعديلات الأخيرة.

الآن أصبح لدينا ملف بيانات في برنامج SPSS باسم "مثال1" يحتوي على ثلاثة متغيرات مُعرّفة بشكل واضح وجاهزة للتحليل الإحصائي، ويكون شكل أيقونة ملف البيانات على سطح المكتب بالصورة .

أما بالنسبة لنافذة المخرجات فيمكن تجاهلها (عدم حفظها) في المرحلة الحالية، وسيتم التعامل معها في أمثلة لاحقة. قم بإغلاق نافذة المخرجات وستظهر لك رسالة للاستفهام عن رغبة المستخدِم في حفظ ملف المخرجات، قم عندها باختيار لا (No) وستُغلق النافذة. ولإنهاء الجلسة الحالية في برنامج SPSS، قم بإغلاق نافذة البرنامج الرئيسية وستظهر لك رسالة تنبيه إلى أن البرنامج سيتم إغلاقه، قم عندها باختيار نعم (Yes).

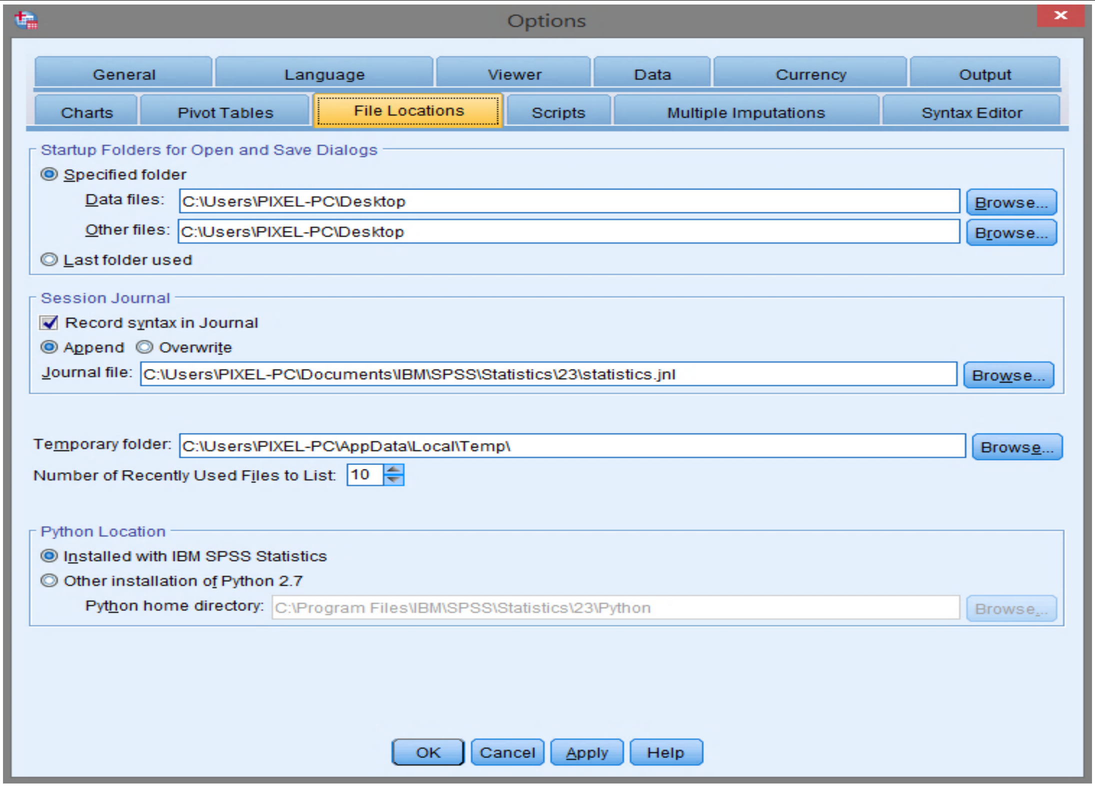
ولفتح ملف البيانات "مثال1" أو أي ملف بيانات آخر، يمكنك إما النقر مرتين على أيقونة الملف في مسار تخزينها، أو فتح برنامج SPSS ثم الضغط[1] على أيقونة فتح الملفات في شريط أيقونات الأوامر ثم اختيار مسار التخزين (وهو سطح المكتب في مثالنا) ثم اختيار الملف المرغوب وهو "مثال1" ثم اختيار فتح (Open). ولاحظ أن نافذة المخرجات ستظهر أيضا مع الملف المفتوح.

5.1 الإعدادات العامة في برنامج SPSS (General Settings)

يمكن في برنامج SPSS، كشأن كل البرامج تحت مظلة ويندوز، تغيير الكثير من الإعدادات الافتراضية في البرنامج، مثل شكل النوافذ، أشكال وألوان الجداول والرسومات البيانية، حجم ونوع ولون الخطوط المستخدمة، كيفية عرض النتائج، وغيرها من الأمور. وهذه الإعدادات يمكن الوصول إليها من شريط الأوامر العلوي باختيار الأمر تحرير (Edit) ثم اختيار خيارات (Options) في آخر القائمة. عندها ستظهر نافذة الإعدادات أو الخيارات، وهنا ننصح المستخدمين المبتدئين بعدم إجراء تغييرات في إعدادات البرنامج قبل أن يكونوا قد قطعوا شوطا جيدا في الالمام بالتعامل مع المخرجات من جداول ورسومات وغيرها.

[1] أو يمكن استخدام الطريقة التقليدية في فتح الملفات عن طريق اختيار أمر ملف (File) في شريط الأوامر ثم فتح (Open) ثم بيانات (Data) ومن ثمة اختيار ملف البيانات من المسار المحدد.

إلا أننا سوف نقوم بتعديل واحد فقط هنا بُغية تسهيل الوصول للملفات المحفوظة. سنقوم بالدخول لنافذة الإعدادات واختيار نافذة **مواقع الملفات** (File Locations) كما يظهر في الشكل (10.1).



شكل 10.1: نافذة الخيارات العامة في برنامج SPSS.

سنقوم في هذه النافذة بتغيير مسار فتح وحفظ ملفات البيانات إلى مسار سطح المكتب (Desktop) بشكل دائم، حيث أنه المسار المفضل للمستخدمين تحت بيئة ويندوز في العموم كما وضحنا سابقا.

في مربع المجلدات الافتتاحية (Startup Folders for Open and Save Dialogs)، قم بالنقر على خيار المجلد المحدد (Specified Folders). الآن سنقوم باختيار مسار سطح المكتب في كلا المربعين؛ ملفات البيانات (Data files) والملفات الأخرى (Other files)، وذلك إما عن طريق كتابة مسار سطح المكتب مباشرة، أو الضغط على زر التصفح (Browse) على اليمين واختيار المسار المطلوب.

ويمكن للقارئ هنا ملاحظة أن مسار سطح المكتب، (في الشكل (10.1))، يحتوي على اسم جهاز الحاسوب الحالي للمؤلف الذي يعمل عليه برنامج SPSS، والذي سيتضمن اسم جهاز المستخدم عند تغيير المسار بالطبع.

الفصل الثاني

التعامل مع ملفات البيانات في برنامج SPSS
(Dealing with Data Files in SPSS)

1.2 استيراد الملفات إلى برنامج SPSS (Importing Files to SPSS)

تناولنا في الفصل السابق طريقة إدخال البيانات في برنامج SPSS بطريقة واجهة المستخدم البيانية ((Graphical User Interface (GUI)) باستخدام نافذة تحرير البيانات، وأوضحنا أن هذه الطريقة تعتبر الطريقة المباشرة والأكثر استخداما في إدخال البيانات في SPSS. إلا أنه في بعض الأحيان، قد تكون البيانات متوفرة مسبقا في ملفات بيانات ضمن برامج أخرى، مثل اكسل (MS Excel)، S-plus، Statistica، R، SAS، وغيرها، عندها وتوفيرا للوقت والجهد يمكن للمستخدِم استيراد هذه الملفات من مصدرها الأصلي إلى برنامج SPSS والتعامل معها مباشرة.

ملاحظة:
في هذا البند، سيتم توضيح كيفية استيراد ملفات البيانات من برنامج اكسل فقط[1] حيث أنه في معظم البرامج الإحصائية الأخرى يمكن بسهولة حفظ أو تصدير ملف البيانات بصيغة اكسل.

لتوضيح طريقة استيراد ملفات البيانات من برنامج اكسل، سنقوم بتطبيق مثال عملي خطوة بخطوة لتسهيل العملية للمستخدِم. قم أولا بإدخال البيانات الموجودة في الجدول التالي، (جدول (1.2))، في ملف بيانات جديد في برنامج اكسل[2]، وقم بحفظه[3] بالاسم "بيانات الطلبة". أي أنه سيأخذ الامتداد ".xlsx".

البيانات في جدول (1.2) هي بيانات افتراضية تمثل المشاهدات الخاصة بـ 35 طالبا جامعيا مقاسة لثمانية متغيرات معرّفة بالصورة التالية.

[1] للمزيد من المعلومات حول استيراد ملفات البيانات بصيغ أخرى غير صيغة اكسل، يمكن الرجوع للمساعدة (Help) في شريط الأوامر الرئيسي في برنامج SPSS.

[2] تم استخدام الإصدارات Excel 2010 و Excel 2016 في التطبيقات وقت إعداد الكتاب.

[3] يُفضل حفظ ملف بيانات اكسل في مسار سطح المكتب لتسهيل استدعاؤه لاحقا.

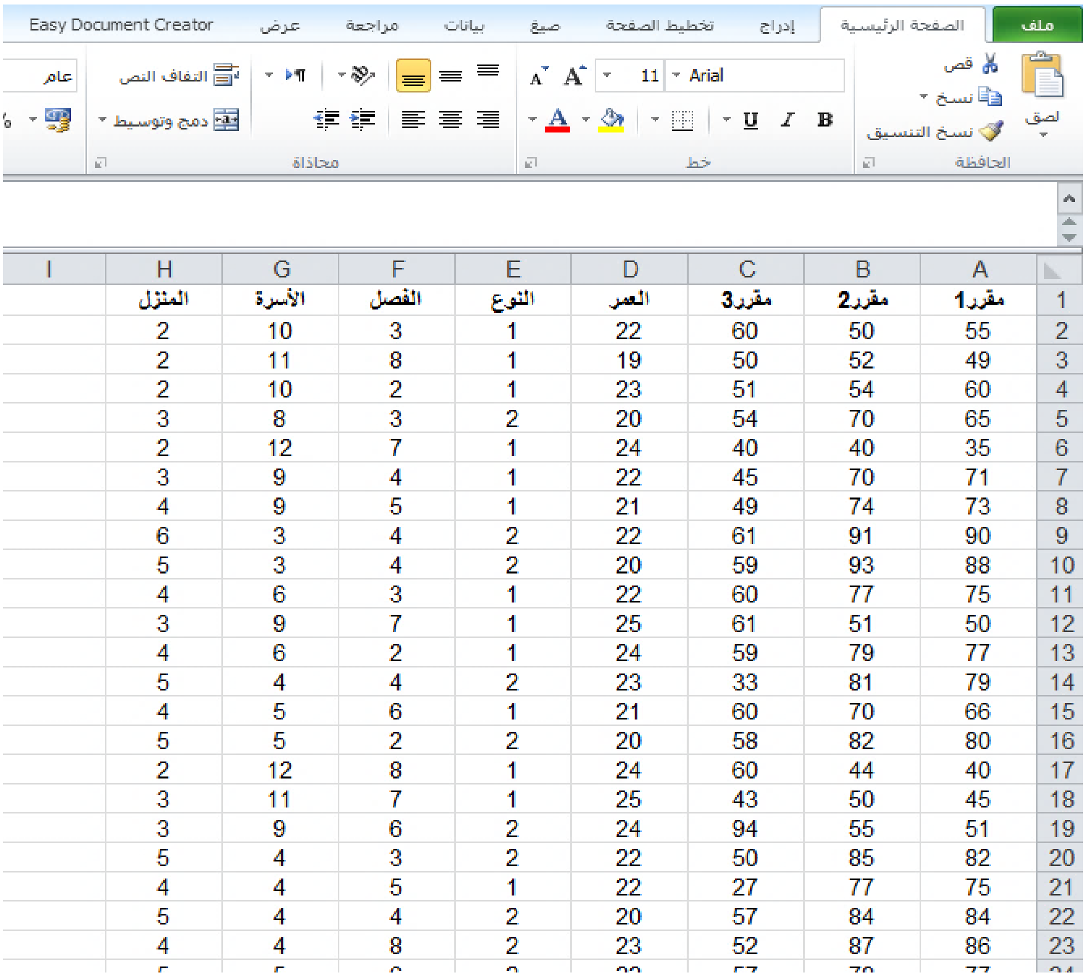
- مقرر1: درجة الطالب[1] في المقرر1.
- مقرر2: درجة الطالب في المقرر2.
- مقرر3: درجة الطالب في المقرر3.
- العمر: عمر الطالب (بالسنوات).
- النوع: جنس الطالب (1=ذكر، و2=أنثى).
- الفصل: ترتيب الفصل الدراسي للطالب.
- الأسرة: عدد أفراد أسرة الطالب.
- المنزل: عدد الحجرات في منزل الطالب.

جدول 1.2: البيانات الخاصة بالملف "بيانات الطلبة".

	مقرر1	مقرر2	مقرر3	العمر	النوع	الفصل	الأسرة	المنزل
1	55	50	60	22	1	3	10	2
2	49	52	50	19	1	8	11	2
3	60	54	51	23	1	2	10	2
4	65	70	54	20	2	3	8	3
5	35	40	40	24	1	7	12	2
6	71	70	45	22	1	4	9	3
7	73	74	49	21	1	5	9	4
8	90	91	61	22	2	4	3	6
9	88	93	59	20	2	4	3	5
10	75	77	60	22	1	3	6	4
11	50	51	61	25	1	7	9	3
12	77	79	59	24	1	2	6	4
13	79	81	33	23	2	4	4	5
14	66	70	60	21	1	6	5	4
15	80	82	58	20	2	2	5	5
16	40	44	60	24	1	8	12	2
17	45	50	43	25	1	7	11	3
18	51	55	94	24	2	6	9	3
19	82	85	50	22	2	3	4	5
20	75	77	27	22	1	5	4	4
21	84	84	57	20	2	4	4	5
22	86	87	52	23	2	8	4	4
23	77	78	57	22	2	6	5	5
24	88	90	62	19	2	7	3	5
25	64	70	50	21	1	7	7	3
26	89	91	53	23	2	4	3	5
27	90	93	50	21	2	6	3	6
28	63	66	51	25	1	5	5	4
29	75	85	95	22	2	3	2	6
30	69	70	44	23	1	5	6	4
31	77	80	92	21	2	8	4	5
32	68	69	54	22	1	6	5	5
33	93	75	45	23	2	5	2	6
34	73	75	52	24	1	6	5	4
35	94	96	49	21	2	4	2	6

[1] درجة الطالب في المقررات الثلاثة مقاسة من 100 درجة.

بعد إدخال بيانات الجدول السابق في اكسل وحفظها بالاسم المطلوب، سنحصل على ملف بيانات مشابه[1] لذلك الموضح في الشكل (1.2).



	A	B	C	D	E	F	G	H
1	مقرر1	مقرر2	مقرر3	العمر	النوع	الفصل	الأسرة	المنزل
2	55	50	60	22	1	3	10	2
3	49	52	50	19	1	8	11	2
4	60	54	51	23	1	2	10	2
5	65	70	54	20	2	3	8	3
6	35	40	40	24	1	7	12	2
7	71	70	45	22	1	4	9	3
8	73	74	49	21	1	5	9	4
9	90	91	61	22	2	4	3	6
10	88	93	59	20	2	4	3	5
11	75	77	60	22	1	3	6	4
12	50	51	61	25	1	7	9	3
13	77	79	59	24	1	2	6	4
14	79	81	33	23	2	4	4	5
15	66	70	60	21	1	6	5	4
16	80	82	58	20	2	2	5	5
17	40	44	60	24	1	8	12	2
18	45	50	43	25	1	7	11	3
19	51	55	94	24	2	6	9	3
20	82	85	50	22	2	3	4	5
21	75	77	27	22	1	5	4	4
22	84	84	57	20	2	4	4	5
23	86	87	52	23	2	8	4	4

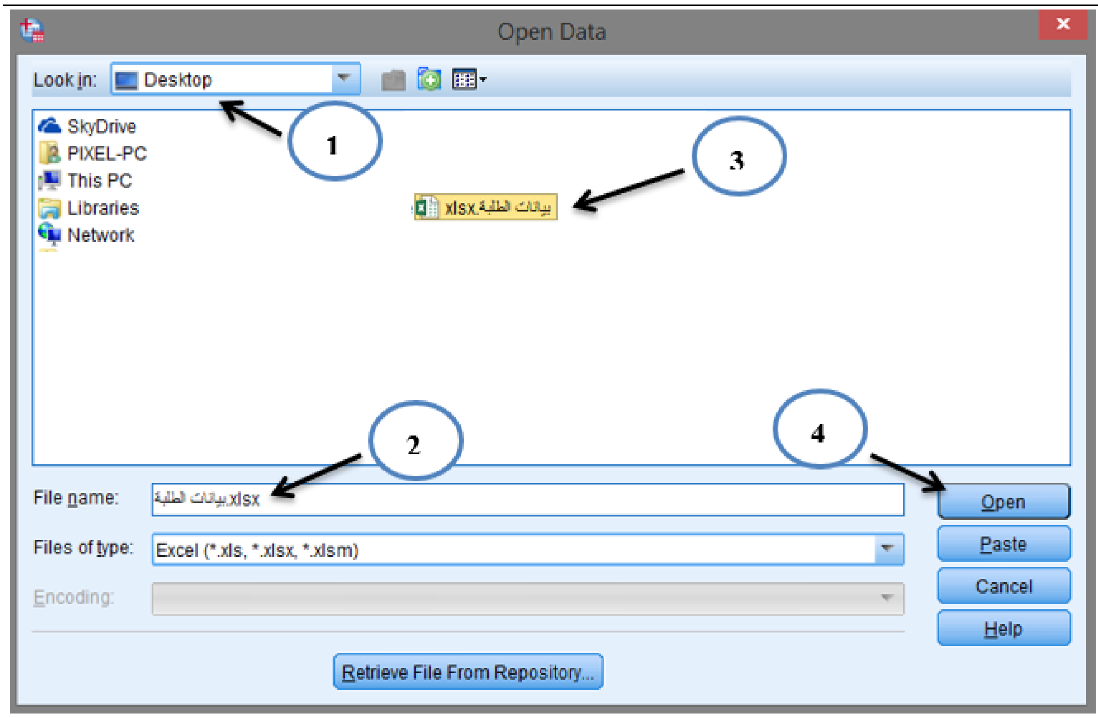
شكل 1.2: ملف البيانات "بيانات الطلبة" في برنامج اكسل.

الآن سنقوم بتنفيذ الخطوات التالية لفتح (استيراد) ملف بيانات بصيغة اكسل في برنامج SPSS:

في برنامج SPSS قم باختيار أيقونة فتح الملفات في شريط أيقونات الأوامر  (أو يمكن اختيار أمر ملف (File) في شريط الأوامر ثم فتح (Open) ثم بيانات (Data)).

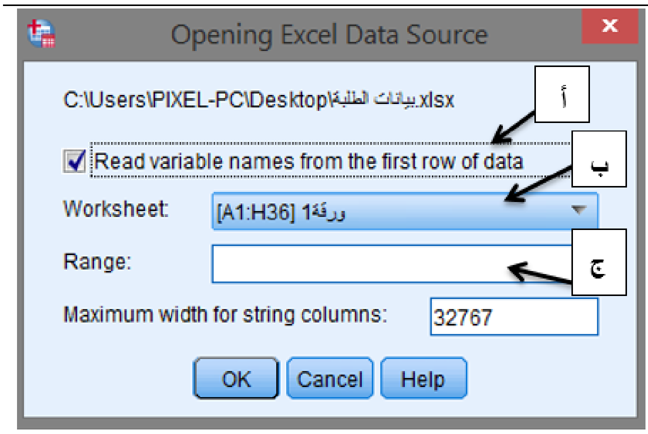
في نافذة فتح الملفات، كما هو موضح في الشكل (2.2)، قم أولا باختيار مسار الملف المطلوب، وهو سطح المكتب الذي تم حفظ الملف فيه، وثانيا قم باختيار نوع الملف المطلوب، وهو ملف اكسل، وسيظهر الملف المطلوب داخل المساحة البيضاء في النافذة، قم باختياره ثم اضغط أمر الفتح (Open).

[1] قد يختلف شكل الملف قليلا إذا ما كان القارئ يستخدم اصدار مختلف عن الإصدارين Excel 2010، أو Excel 2016، أو إذا كان يستخدم الواجهة الإنجليزية للبرنامج.



شكل 2.2: خطوات فتح ملف اكسل في نافذة فتح الملفات في برنامج SPSS.

بعد الضغط على أمر الفتح، ستظهر نافذة جديدة، كما هو ظاهر في الشكل (3.2)، والتي تحتوي على بعض الخيارات التي يمكن للمستخدِم الاستفادة منها كالتالي؛



شكل 3.2: الخيارات الخاصة بفتح ملف اكسل في برنامج SPSS.

الخيار (أ) يتم اختياره عندما تكون أسماء المتغيرات في ملف البيانات موجودة في الصف الأول، (وهذا هو الحال في مثالنا الحالي). أما الخيار (ب)، فهو لتحديد الورقة التي توجد فيها البيانات في ملف اكسل الأصلي، (وفي مثالنا البيانات موجودة في الورقة[1] الأولى؛ ورقة1). وفي الخيار الثالث (ج)، يمكن للمستخدِم اختيار مجموعة

[1] حيث أن برنامج اكسل يوفر افتراضيا ثلاثة ورقات لكل ملف بيانات.

محددة من المتغيرات التي يرغب باستيرادها من داخل ملف اكسل إذا لم يرغب باستيراد كل المتغيرات، (وفي مثالنا تم ترك هذا الخيار فارغا بمعنى أن كل المتغيرات في الملف سيتم استيرادها).

بعد الانتهاء من خيارات الفتح اضغط موافق (OK)، وسيتم فتح ملف البيانات المطلوب، وكذلك فتح نافذة المخرجات كالمعتاد، وستجد أن كافة المتغيرات قد تم استيرادها بنفس الأسماء إلى برنامج SPSS. وبالطبع، سنقوم بحفظها بصيغة SPSS، قبل الانتقال للخطوة التالية. وخطوات الحفظ ستكون مطابقة لما تم توضيحه في الفصل الأول (في البند (2.4.1))، أي أن المستخدِم سيقوم باختيار File>Save ثم اختيار مسار الحفظ، (وليكن كما هو الحال دائما على سطح المكتب)، ثم اختيار اسم الملف، (وليكن نفس الاسم الأصلي المستخدَم "بيانات الطلبة")، ثم نضغط حفظ (Save).

<u>ملاحظة:</u> لاحظ هنا أنه سيكون لديك الآن ملفان مختلفان بنفس الاسم على سطح المكتب، الأول هو بصيغة اكسل بالامتداد "بيانات الطلبة.xlsx"، والآخر بصيغة SPSS وهو بالامتداد "بيانات الطلبة.sav".

الآن في ملف بيانات SPSS؛ "بيانات الطلبة.sav" يمكنك ملاحظة أن خصائص المتغيرات بحاجة إلى بعض التعديل، وفيما يلي سنقوم بإجراء التغيرات التالية بهدف جعل الملف أكثر تنظيما للاستخدامات اللاحقة؛

- تغيير مسافة عرض قيم المتغير (Width) لجميع المتغيرات إلى القيمة الافتراضية 8.
- تغيير عدد الخانات العشرية (Decimals) لجميع المتغيرات إلى 0 نظرا لطبيعة الأعداد الصحيحة للمتغيرات.
- إضافة الأوصاف التالية للمتغيرات (Label)؛
  1. مقرر1: درجة الطالب في المقرر1.
  2. مقرر2: درجة الطالب في المقرر2.
  3. مقرر3: درجة الطالب في المقرر3.
  4. العمر: عمر الطالب بالسنوات.
  5. النوع: جنس الطالب.
  6. الفصل: ترتيب الفصل الدراسي للطالب.
  7. الأسرة: عدد أفراد أسرة الطالب.
  8. المنزل: عدد الحجرات في منزل الطالب.
- إضافة التصنيف (Values) التالي لقيم متغير النوع؛ (1= ذكر، و2= أنثى).
- تغيير عرض العامود (Columns) لجميع المتغيرات إلى القيمة الافتراضية 8.

- إبقاء أنواع المتغيرات (Measure)؛ "مقرر1"، "مقرر2"، "مقرر3"، و"العمر" كمية (Scale)، ومتغير "النوع" وصفي اسمي (Nominal)، وتغيير نوع المتغير؛ "الفصل" إلى وصفي ترتيبي (Ordinal)، وتغير نوع المتغيران؛ "الأسرة" و"المنزل" إلى كمي (Scale).

نقوم بعد ذلك بحفظ ملف البيانات، والذي سيبدو كما هو موجود في الشكل (4.2). ثم نغلق الملف.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
مقرر1	Numeric	8	0	درجة الطالب في المقرر1	None	None	8	Right	Scale	Input
مقرر2	Numeric	8	0	درجة الطالب في المقرر2	None	None	8	Right	Scale	Input
مقرر3	Numeric	8	0	درجة الطالب في المقرر3	None	None	8	Right	Scale	Input
العمر	Numeric	8	0	عمر الطالب بالسنوات	None	None	8	Right	Scale	Input
النوع	Numeric	8	0	جنس الطالب	{1, ذكر}...	None	8	Right	Nominal	Input
الفصل	Numeric	8	0	ترتيب الفصل الدراسي للطالب	None	None	8	Right	Ordinal	Input
الأسرة	Numeric	8	0	عدد أفراد أسرة الطالب	None	None	8	Right	Scale	Input
المنزل	Numeric	8	0	عدد الحجرات في منزل الطالب	None	None	8	Right	Scale	Input

بيانات الطلبة.sav [DataSet1] - IBM SPSS Statistics Data Editor

View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

مقرر1	مقرر2	مقرر3	العمر	النوع	الفصل	الأسرة	المنزل
55	50	60	22	1	3	10	2
49	52	50	19	1	8	11	2
60	54	51	23	1	2	10	2
65	70	54	20	2	3	8	3
35	40	40	24	1	7	12	2
71	70	45	22	1	4	9	3
73	74	49	21	1	5	9	4
90	91	61	22	2	4	3	6
88	93	59	20	2	4	3	5
75	77	60	22	1	3	6	4
50	51	61	25	1	7	9	3
77	79	59	24	1	2	6	4

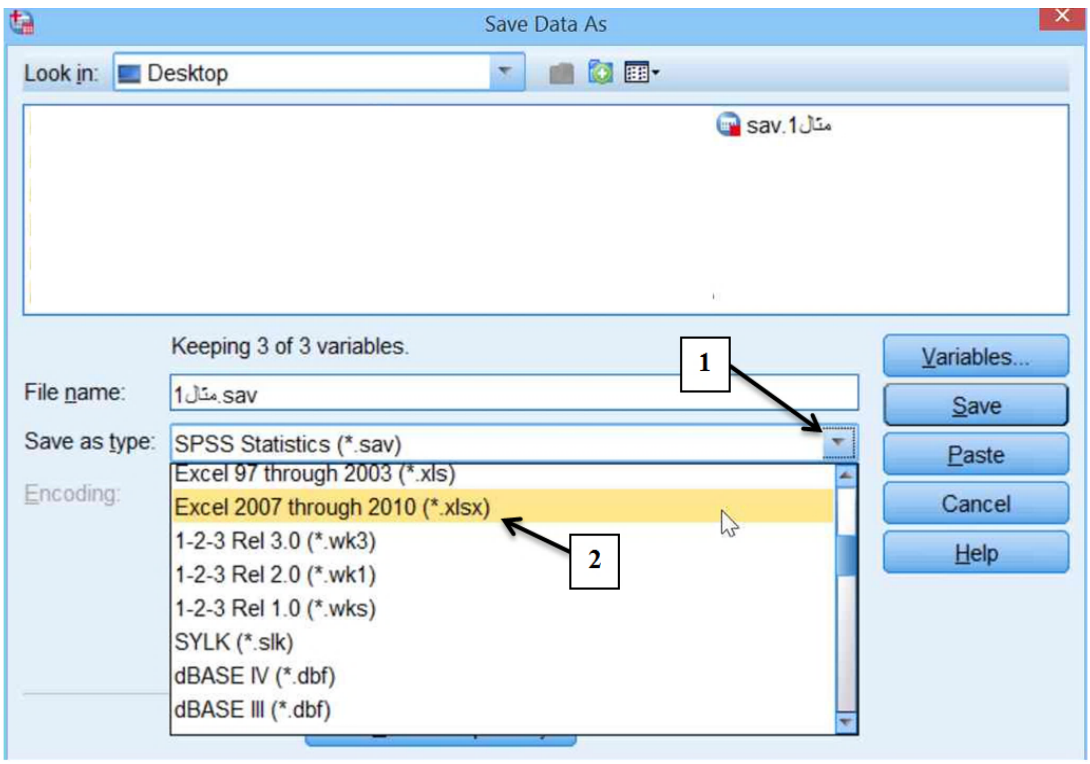
شكل 4.2: أسلوبي عرض المتغيرات (في الأعلى) وعرض البيانات (في الأسفل) لملف البيانات "بيانات الطلبة" في **SPSS**.

2.2 تصدير الملفات من برنامج SPSS (Exporting Files from SPSS)

كما هو الحال مع استيراد ملفات البيانات من البرامج الأخرى، (والذي تم تطبيقه مع برنامج اكسل)، فإنه يمكن وبصورة أبسط، تصدير الملفات المحفوظة بصيغة SPSS إلى البرامج الأخرى. وما سنقوم بتطبيقه في هذا البند، هو حفظ ملف البيانات (المتوفر مسبقا بصيغة SPSS) بصيغة أخرى، وهي صيغة برنامج اكسل باعتباره "الصيغة المشتركة" بين معظم البرامج الإحصائية والرياضية. ومن أهم البرامج التي يمكن للمستخدِم حفظ ملف البيانات بصيغتها المباشرة (بإصدارات مختلفة) هي البرامج التالية؛ اكسل، dBase، SAS، Stata، وكذلك يمكن حفظ ملف البيانات بصورة نصية (بالامتداد .txt).

ولنأخذ ملف البيانات الذي قمنا بتكوينه في الفصل الأول باسم "مثال1" في برنامج SPSS كمثال تطبيقي على حفظ الملفات بصيغة اكسل. وستكون خطوات تصدير الملف كالتالي:

1. قم بفتح ملف البيانات "مثال1" الموجود في مسار سطح المكتب في برنامج SPSS.
2. في شريط الأوامر العلوي، قم باختيار ملف (File) ثم حفظ باسم (Save As)، فتظهر النافذة الخاصة بالحفظ، كما يظهر في الشكل (5.2).
3. في هذه النافذة، قم بالضغط على الخيار الخاص بتحديد نوع الملف المطلوب حفظه، كما يوضح السهم رقم (1) الموجود في الشكل. ستنسدل عند ذلك قائمة صغيرة تحتوي على أسماء البرامج التي يمكن اختيار صيغة (وإصدار) حفظ الملف بها. في مثالنا، سنقوم بحفظ الملف بصيغة اكسل (الإصدار 2010) كما يشير السهم (2) في الشكل.
4. بعد ذلك قم بالضغط على أمر حفظ (Save) في النافذة. وبذلك تكون عملية حفظ ملف البيانات (بنفس الاسم "مثال1") بصيغة اكسل قد تمت، ويمكنك بعدها مشاهدة ملف البيانات "مثال1.xlsx" على سطح المكتب. قم بفتحه وستجد البيانات المتعلقة بالمتغيرات الثلاثة؛ الوزن، العمر، والنوع موجودة فيه بنفس النسق.



شكل 5.2: خطوات حفظ ملف البيانات "مثال1" بصيغة اكسل بداخل برنامج SPSS.

3.2 معالجة البيانات (Data Manipulation)

في كثير، إن لم يكن معظم الأحيان، قد يحتاج المستخدِم أو الباحث إلى إعادة ترتيب أو معالجة بعض المتغيرات داخل ملف البيانات بناء على تقسيم معين[1]، أو قد تتطلب الدراسة الإحصائية تعريف متغير جديد أو أكثر اعتمادا على علاقته[2] مع متغير أو أكثر في البيانات الأصلية، أو ربما تعريف متغير جديد كدالة[3] رياضية في متغير آخر، وغيرها من الحالات التي تستدعي معالجة البيانات الأصلية لتحقيق أهداف معينة في الدراسة أو التحليل الإحصائي، في هذه الحالة، يمكن استخدام الأدوات الخاصة بمعالجة أو ترتيب البيانات.

ويمكن أيضا استخدام أداة ترتيب المشاهدات للمساعدة في إعادة ترتيب القيم تصاعديا أو تنازليا إما لتنظيمها بصورة أكثر وضوحا، أو لتقسيم البيانات المرتبة استعدادا لاستخدامها في تحليل البيانات لاحقا. ويمكن في برنامج SPSS تقسيم البيانات بناء على قيم متغير واحد أو متغيرين أو حتى أكثر من ذلك.

ولتوضيح كيفية تنفيذ ترتيب البيانات في SPSS، لنأخذ مثالا عمليا لترتيب البيانات الخاصة بملف البيانات "مثال1" الذي تم إنشاؤه في الفصل الأول. قم بفتح الملف من مسار سطح المكتب ثم اختر أسلوب عرض البيانات (Data View)، فتظهر البيانات، (كما هو موضح في الشكل (9.1) في الفصل الأول).

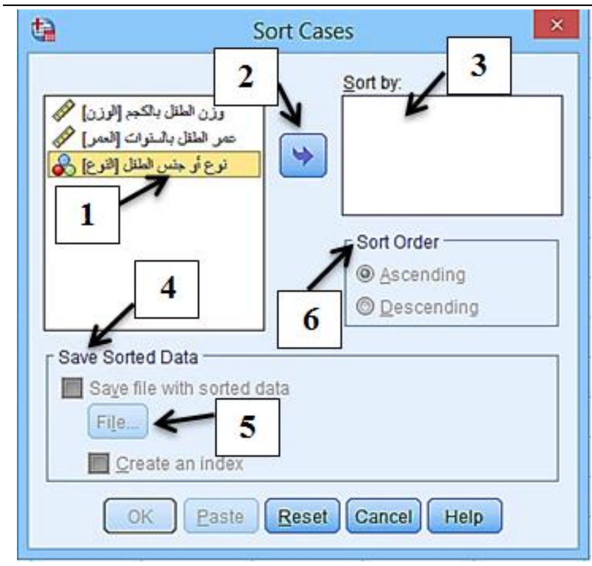
سنقوم أولا في هذا المثال بترتيب أوزان الأطفال وأعمارهم بحسب جنس الطفل، أي أن متغيري "الوزن" و"العمر" سيتم إعادة ترتيب قيمهما بحسب جنس الطفل ذكرا كان أو أنثى. وحيث أن الذكور في هذه البيانات تم تعيينهم للقيمة 1 والإناث للقيمة 2، فإن أوزان وأعمار الأطفال الذكور سيتم عرضها في الصفوف الأولى متبوعة بأوزان وأعمار الإناث. وسنقوم الآن بتنفيذ الخطوات التالية:

1. في شريط الأوامر العلوي، قم باختيار بيانات (Data) ثم اختار ترتيب المشاهدات (Sort Cases) من القائمة المنسدلة، فتظهر نافذة جديدة كما هو موضح في الشكل (6.2).
2. وحيث أننا سنقوم بتقسيم متغيري "الوزن" و"العمر" بحسب متغير نوع الطالب، قم باختيار متغير "النوع" كما هو مشار إليه في رقم (1) في الشكل. ثم اضغط على السهم المشار إليه في رقم (2)، فينتقل المتغير إلى الخانة المشار إليها بالرقم (3)، والتي تشير لمتغير التصنيف المحدد.

[1] كأن يتم ترتيب المتغيرات بناء على النوع أو الوزن أو المستوى التعليمي، وغيرها.

[2] مثل أن يتم تعريف متغير جديد يمثل مدخرات الأشخاص بناء على الراتب الشهري والدخل الإضافي والمصروف الشهري.

[3] مثل تعريف متغير كتلة الجسم كدالة في متغير الوزن.



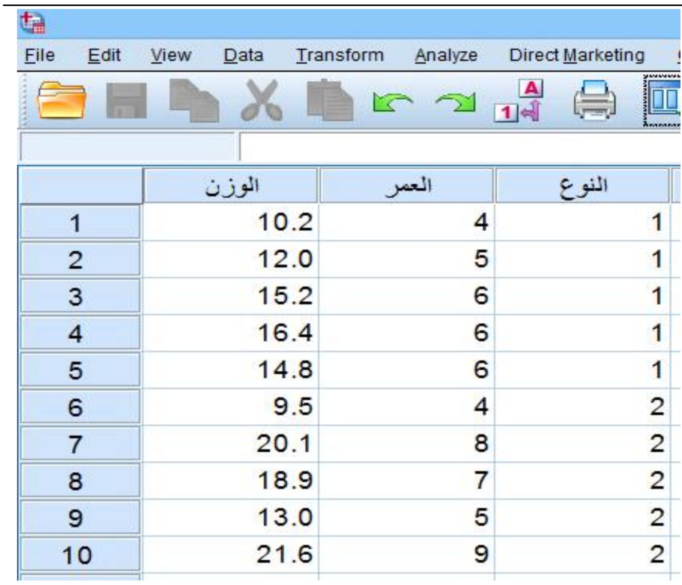
شكل 6.2: نافذة ترتيب المشاهدات لملف البيانات "مثال1".

ملاحظة:

إذا ما قمت بالضغط على (OK) في النافذة، فسيتم تنفيذ إعادة الترتيب المطلوب في نفس ملف البيانات الأصلي "مثال1" وهذا سيؤدي بدوره لتغيير ترتيب البيانات بصورة نهائية، لذلك يُفضل عادة حفظ التغييرات الجديدة في ملف بيانات جديد لتمكين المستخدم من العودة للبيانات الأصلية متى أراد ذلك. وهذا ما سنقوم به في الخطوة التالية.

3. قم باختيار الخيار المشار إليه في الرقم (4) في شكل (6.2)، ثم اضغط على أمر ملف (File) المشار إليه بالرقم (5) وذلك لتحديد مسار واسم الملف الجديد الذي سيحتوي على البيانات المُعاد ترتيبها. ستظهر بالطبع نافذة حفظ البيانات المعتادة، قم باختيار مسار سطح المكتب ثم قم بتعيين الاسم "مثال2" للملف الجديد واضغط حفظ (Save). ستلاحظ بعد ذلك، في نفس النافذة، ظهور مسار الملف واسمه بجانب أمر ملف (File).
4. كخطوة اختيارية، (مشار إليها بالرقم (6) في الشكل (6.2))، يمكن للمستخدِم ترتيب متغيري الوزن والعمر بحسب الترتيب التصاعدي (Ascending) أو التنازلي (Descending) لمتغير النوع، أي يمكن أن يكون التصنيف الأول هو للذكور المعرفين بالقيمة 1 (تصاعديا) أو للإناث المعرفات بالقيمة 2 (تنازليا). في مثالنا، سنبقي الترتيب الافتراضي، وهو التصاعدي، كما هو بدون تغيير، وسننهي خطوات إعادة ترتيب المتغيرات بالضغط على موافق في النافذة، فيظهر[1] ملف البيانات الجديد بالاسم "مثال2"، كما يظهر في الشكل (7.2). ولاحظ من الشكل كيف أن قيم مشاهدات الوزن والعمر للأطفال الذكور قد ظهرت في الصفوف الأولى متبوعة بالقيم الخاصة بالأطفال الإناث.

[1] لاحظ أن ملف البيانات الأصلي "مثال1" سيُغلق تلقائيا.

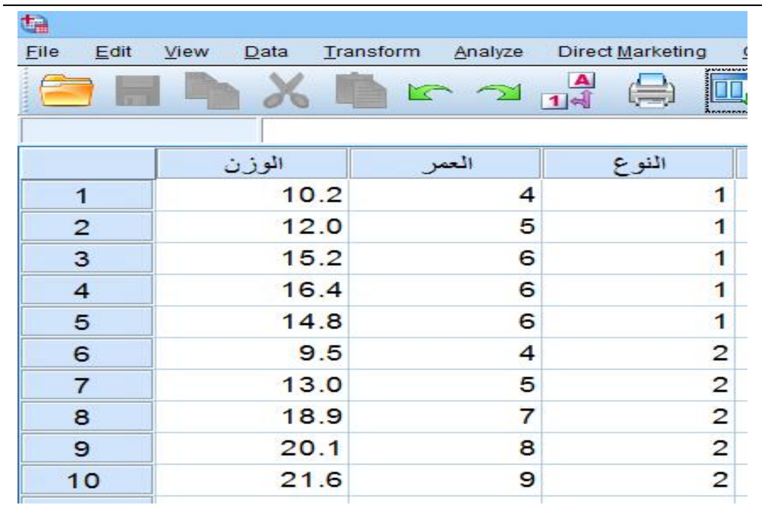


	الوزن	العمر	النوع
1	10.2	4	1
2	12.0	5	1
3	15.2	6	1
4	16.4	6	1
5	14.8	6	1
6	9.5	4	2
7	20.1	8	2
8	18.9	7	2
9	13.0	5	2
10	21.6	9	2

شكل 7.2: البيانات المصنفة بحسب جنس الطالب، (ملف البيانات "مثال2").

ويمكن إعادة ترتيب البيانات بناء على <u>متغيرين</u> بدلا من متغير واحد، كما أشرنا سابقا. ولتنفيذ ذلك عمليا، افترض أننا نريد ترتيب البيانات في الملف "مثال1" بناء على جنس الطفل أولا ثم عمر الطفل ثانيا، فنقوم بفتح[1] ملف البيانات "مثال1" من جديد، ثم إعادة تنفيذ الخطوات السابقة مع إجراء التغييرات التالية؛

- في الخطوة الثانية: وبعد اختيار متغير النوع ونقله للخانة المشار إليها بالرقم (3)، قم باختيار متغير العمر ثم نقله للخانة (3).
- في الخطوة الثالثة: قم بتعيين الاسم "مثال3" للملف الجديد واضغط حفظ (Save)، ثم اضغط موافق (OK) فيظهر ملف البيانات المطلوب كما يظهر في الشكل (8.2).



	الوزن	العمر	النوع
1	10.2	4	1
2	12.0	5	1
3	15.2	6	1
4	16.4	6	1
5	14.8	6	1
6	9.5	4	2
7	13.0	5	2
8	18.9	7	2
9	20.1	8	2
10	21.6	9	2

شكل 8.2: البيانات المصنفة بحسب جنس الطالب ثم عمره، (ملف البيانات "مثال3").

[1] لاحظ هنا أن ملف البيانات "مثال2" سيبقى مفتوحا.

ولاحظ هنا الفرق في طريقة عرض البيانات في الملفين أن البيانات في الملف "مثال3" هي مرتبة بحسب جنس الطفل أولا (ذكور ثم إناث)، ويلي ذلك الترتيب الثاني بحسب عمر الطفل تصاعديا ضمن كل جنس، بمعنى أن أوزان الأطفال ستكون موزعة أو مصنفة إلى قسمين؛ الذكور ثم الإناث، وأوزان الذكور (المناظرة للقيمة 1 في متغير النوع) ستكون موزعة بحسب أعمارهم (بداية من 4 إلى 6 سنوات)، وأوزان الإناث أيضا (المناظرة للقيمة 2 في متغير النوع) ستكون موزعة بحسب أعمارهن (بداية من 4 إلى 9 سنوات).

وهكذا، فإنه يمكن إعادة ترتيب أو تصنيف أي متغير أو أكثر بناء على متغير آخر أو أكثر بنفس الطريقة السابقة. ويمكن للقارئ، كتمرين إضافي، إعادة ترتيب بيانات المثال "مثال1" بحسب العمر أو الوزن فقط، أو بحسب العمر ثم الجنس، ...، وهكذا، وتسمية البيانات الجديدة بأسماء مختلفة.

4.2 حساب تكرار المشاهدات (Counting Case Occurrence)

في كثير من الدراسات التي يعتمد جمع البيانات فيها على توزيع استبيان (Questionnaire) على مجموعة من الأشخاص أو الهيئات وأخذ آراؤهم في مواضيع محددة، قد يحتاج الباحث إلى حساب (أي جمع) رأي كل مستبْين (مستهدف) لمجموعة من الأسئلة للخروج بقيمة عددية (كمية) تعكس مدى موافقة أو عدم موافقة المستبْين على هذه الأسئلة، وفي هذه الحالة يمكن باستخدام برنامج SPSS حساب تكرار هذه الآراء وتعريفها كمتغير جديد يتم إضافته للبيانات الأصلية. وسنقوم فيما يلي بتوضيح كيفية تنفيذ هذه المهمة باستخدام مثال توضيحي.

البيانات في الجدول (2.2) تمثل آراء أو أجوبة عينة مكونة من 12 شخص ردا على المطلوب التالي: "ضع إشارة (√) أمام التطبيقات التي تستخدمها وإشارة (×) أمام التطبيقات التي لا تستخدمها". وهذه التطبيقات هي أشهر أربعة تطبيقات في مجال التواصل الاجتماعي؛ فيسبوك (Facebook)، واتساب (WhatsApp)، فايبر (Viber)، وإنستجرام (Instagram). وخلال مرحلة تفريغ الاستبيان، تم إدراج القيم (م1، م2، ...، م12) كمتغير يمثل أسماء الأشخاص في العينة، وتم تعيين القيمة 1 للإجابة بنعم (√) وتعيين القيمة 0 للإجابة بلا (×) وذلك لكل تطبيق من التطبيقات الأربعة. فمثلا، الشخص الأول (م1) كانت إجابته بأنه يستخدم تطبيقي فيسبوك وواتساب ولا يستخدم تطبيقي فايبر وإنستجرام، وهكذا.

جدول 2.2: الآراء الخاصة بعينة من الأشخاص حول استخدام بعض تطبيقات التواصل الاجتماعي.

	المستخدم	فيسبوك	واتساب	فايبر	إنستجرام
1	م1	1	1	0	0
2	م2	1	0	0	0
3	م3	0	1	0	0
4	م4	1	1	1	1
5	م5	0	0	1	0
6	م6	0	0	0	0
7	م7	1	1	1	0
8	م8	1	0	1	1
9	م9	1	0	0	1
10	م10	1	1	1	1
11	م11	1	0	1	1
12	م12	0	1	1	0

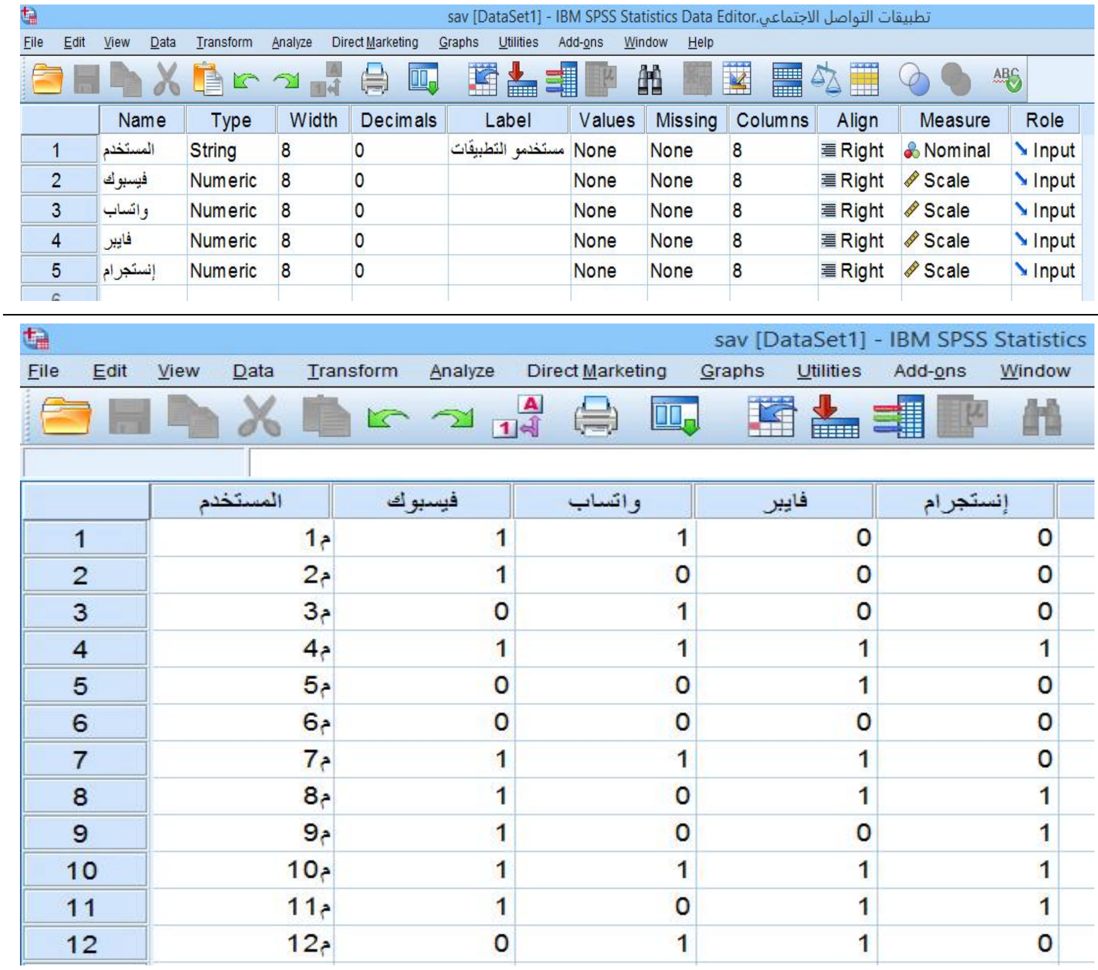
وهذا النوع من الترميز للمتغيرات يندرج تحت ما يسمى **بالمتغير الوهمي** (Dummy Variable)، والذي سنتناوله بشرح أكثر تفصيلا في الفصل الثالث، (في البند 3.3.3 تحديدا).

قم أولا بإدخال بيانات الجدول (2.2) في ملف بيانات جديد في SPSS باسم "تطبيقات التواصل الاجتماعي"، ثم قم بإجراء التغييرات الخاصة كما هو موضح في الشكل التالي (شكل (9.2)).

ملاحظة:

بالنسبة للمتغير الأول؛ "المستخدم"، قم بتغيير طبيعة المتغير (Type) إلى (String) بدلا من (Numeric) لكي تتمكن من إدخال أسماء المستخدمين كقيم غير عددية للمتغير.

المطلوب تنفيذه في هذا المثال هو جمع كل الإجابات بنعم لمستخدمي التطبيقات، أي تعريف متغير جديد يمثل عدد التطبيقات التي يستخدمها كل شخص من بين هذه التطبيقات الأربعة، وبمعنى أبسط؛ المتغير الجديد سيكون عبارة عن حاصل جمع كل صف في الأعمدة الأربعة الأخيرة.



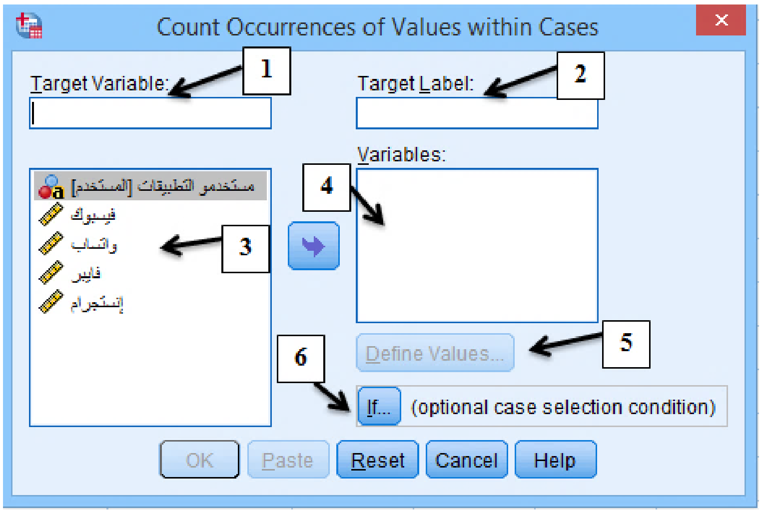
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	المستخدم	String	8	0	مستخدم التطبيقات	None	None	8	Right	Nominal	Input
2	فيسبوك	Numeric	8	0		None	None	8	Right	Scale	Input
3	واتساب	Numeric	8	0		None	None	8	Right	Scale	Input
4	فايبر	Numeric	8	0		None	None	8	Right	Scale	Input
5	إنستجرام	Numeric	8	0		None	None	8	Right	Scale	Input

	المستخدم	فيسبوك	واتساب	فايبر	إنستجرام
1	م1	1	1	0	0
2	م2	1	0	0	0
3	م3	0	1	0	0
4	م4	1	1	1	1
5	م5	0	0	1	0
6	م6	0	0	0	0
7	م7	1	1	1	0
8	م8	1	0	1	1
9	م9	1	0	0	1
10	م10	1	1	1	1
11	م11	1	0	1	1
12	م12	0	1	1	0

شكل 9.2: أسلوبي عرض المتغيرات (في الأعلى) وعرض البيانات (في الأسفل) لملف البيانات "تطبيقات التواصل الاجتماعي".

في نافذة عرض البيانات لملف "تطبيقات التواصل الاجتماعي"، قم باختيار أمر **تحويل** (Transform) في شريط الأوامر العلوي، ثم اختر أمر **عَد القيم ضمن المشاهدات** (Count Values within Cases) من القائمة المنسدلة. فتظهر لك النافذة الخاصة بتنفيذ هذه العملية كما يُشاهد في الشكل (10.2).

في الخانة (1)، وهو مربع المتغير المستهدَف (Target Variable)، يتم إدخال اسم للمتغير الجديد الذي سيمثل تكرارات القيمة المطلوبة (وهي القيمة 1 في مثالنا)، وليكن هذا الاسم هو "التطبيقات". وفي الخانة (2)، وهو وصف المتغير المستهدَف (Target Label)، يتم كتابة وصف لما يمثله هذا المتغير الجديد، وليكن الوصف هو "عدد التطبيقات المستخدمة".



شكل 10.2: النافذة الخاصة بتنفيذ عد تكرارات القيم ضمن المشاهدات.

بعد ذلك، وفي الخانة (3)، يتم اختيار [1] المتغيرات التي يرغب المستخدم في جمع تكرارات قيمها، وفي مثالنا ستكون المتغيرات المطلوبة هي "فيسبوك"، "واتساب"، "فايبر"، و"إنستجرام". وبعد اختيارها، اضغط على سهم نقل المتغيرات فتنتقل المتغيرات التي تم اختيارها إلى مربع المتغيرات (Variables) في الخانة (4).

المرحلة التالية ستكون تحديد القيمة المطلوب جمع تكراراتها، (وهي القيمة 1 في كل المتغيرات الأربعة)، ويتم ذلك عن طريق الضغط على خيار **تعريف القيم** (Define Values) المشار إليه بالسهم (5). فتظهر نافذة جديدة كما يظهر في الشكل التالي، (شكل (11.2)).

ملاحظة:

بالنسبة للخيار (6)، وهو الدالة الشرطية **إذا** (If)، فيمكن استخدامه إذا ما أراد المستخدم وضع قيد أو شرط على عملية عد التكرارات ضمن المتغيرات بحيث يتم اعتبار أو استثناء قيم محددة بناء على الشرط أو الشروط الموضوعة من قِبل المستخدم.

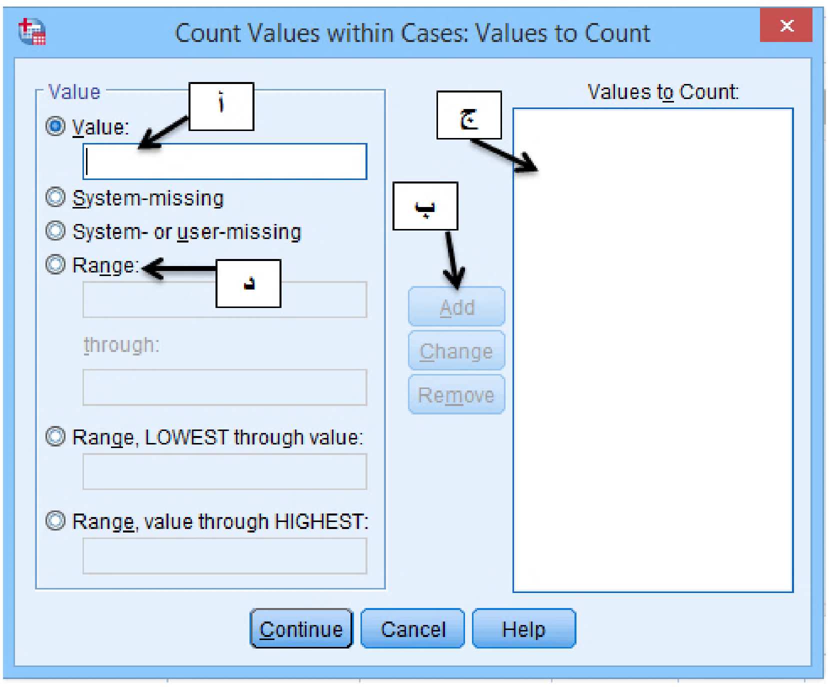
[1] يمكن أن يتم اختيار كل متغير على حدة أو تظليل كل المتغيرات المطلوبة واختيارها معا.

في هذه النافذة الجديدة، (شكل (11.2))، يتم إدخال القيمة المطلوب عد تكراراتها في المتغيرات في الخانة (أ)، وهو مربع القيمة (Value)، وفي مثالنا سيتم إدخال القيمة 1 في هذه الخانة.

وبمجرد كتابة[1] القيمة 1، سيضاء مربع خيار **إضافة** (Add) المشار إليه بالسهم (ب)، قم بالضغط عليه فتنتقل القيمة 1 إلى الخانة (ج)، وهو مربع القيم التي سيتم عدها (Values to count). قم بعدها بإنهاء العملية عن طريق الضغط على استمرار (Continue) في أسفل النافذة.

<u>**ملاحظة:**</u>

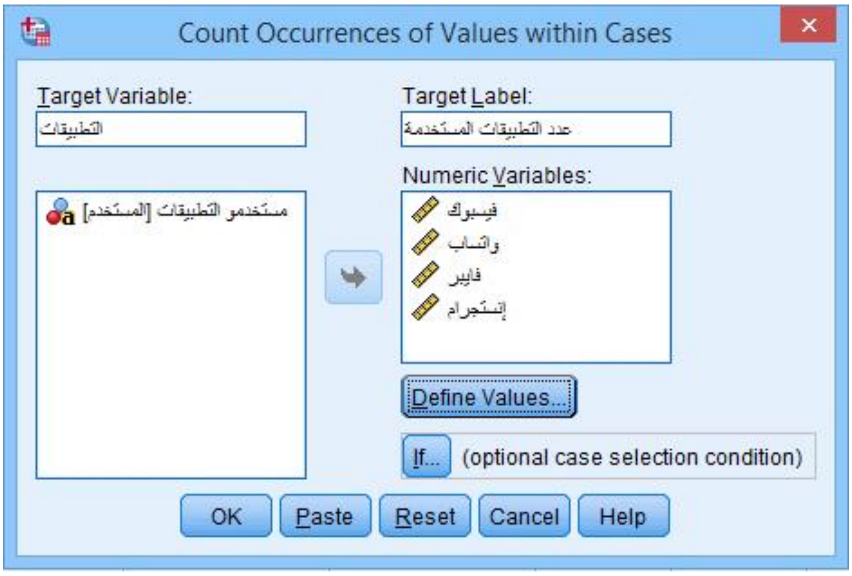
بالنسبة للخيار (د)، وهو **المدى** (Range)، فيُستخدَم عندما نرغب بِعَد تكرارات قيم تتراوح في مدى أو فترة معينة وليس عَد قيم مفردة، فمثلا إذا كان المطلوب هو عد تكرارات قيم من 2 إلى 4 فقط من ضمن متغيرات معينة تضم قيما من 1 إلى 5، فيتم اختيار الخيار (د) ثم كتابة 2 في الخانة الأولى التي تليه وكتابة 4 في الخانة الثانية. أما الخياران الأخيران؛ (Range, LOWEST through value) و(Range, value through HIGHEST) فيُستخدمان عندما يكون المطلوب هو عد تكرارات قيم من قيمة معينة فأعلى منها، أو قيمة معينة فأدنى منها، على الترتيب.



شكل 11.2: النافذة الخاصة بتحديد قيم التكرارات المطلوب عدّها ضمن المشاهدات.

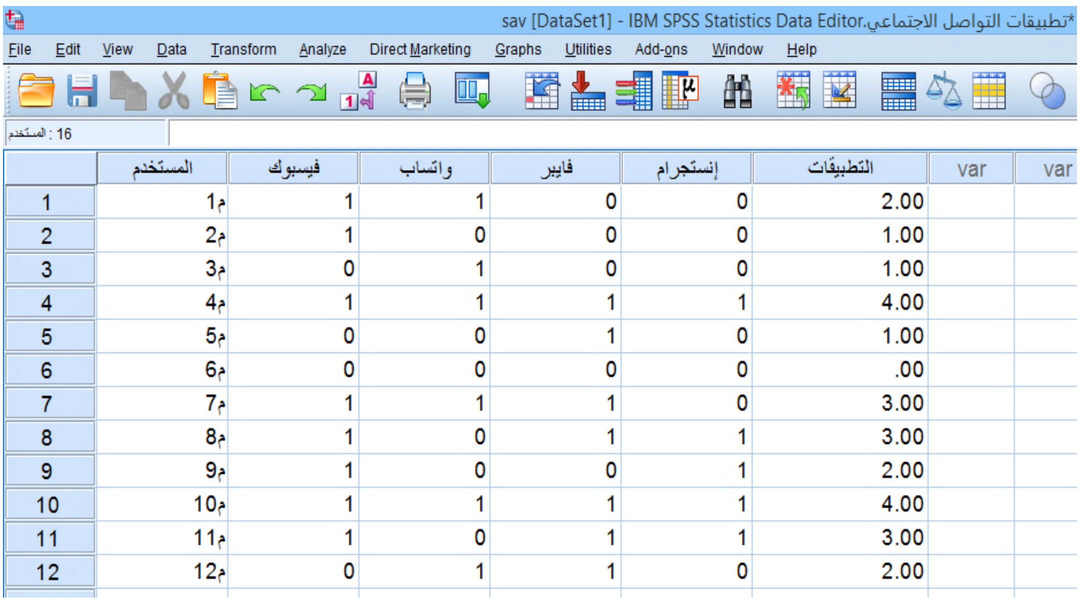
[1] بدون الضغط على أمر الإدخال (Enter) في لوحة المفاتيح.

بعد الضغط على استمرار (Continue) في النافذة السابقة، سيتم العودة للنافذة الأولى، (الشكل (10.2))، والتي ستبدو بالصورة التالية، (شكل (12.2))؛



شكل 12.2: النافذة الخاصة بتنفيذ عد تكرارات القيم ضمن المشاهدات بعد تطبيق الخيارات.

اضغط موافق (OK) لإنهاء العملية بالكامل، وستلاحظ ظهور متغير جديد، هو "التطبيقات"، كما يُشاهد في الشكل (13.2)، والذي يمثل عدد التطبيقات التي يستخدمها كل شخص في عينة البيانات من بين التطبيقات الأربعة. (ويمكن للقارئ تغيير عدد الخانات العشرية للمتغير الجديد إلى 0، للحصول على مظهر أفضل للبيانات).



	المستخدم	فيسبوك	واتساب	فايبر	إنستجرام	التطبيقات	var	var
1	م1	1	1	0	0	2.00		
2	م2	1	0	0	0	1.00		
3	م3	0	1	0	0	1.00		
4	م4	1	1	1	1	4.00		
5	م5	0	0	1	0	1.00		
6	م6	0	0	0	0	.00		
7	م7	1	1	1	0	3.00		
8	م8	1	0	1	1	3.00		
9	م9	1	0	0	1	2.00		
10	م10	1	1	1	1	4.00		
11	م11	1	0	1	1	3.00		
12	م12	0	1	1	0	2.00		

شكل 13.2: عرض بيانات ملف "تطبيقات التواصل الاجتماعي" بعد تعريف متغير "التطبيقات".

5.2 إعادة ترميز المتغيرات (Recoding Variables)

يوفر برنامج SPSS أيضا إمكانية إعادة ترميز أو تغيير قيم متغير أو أكثر في البيانات بحسب ما تتطلبه الدراسة أو التحليل الإحصائي. ويُقصد بإعادة الترميز هنا تغيير بعض أو كل قيم المتغير بشكل تلقائي وليس المقصود هو تصحيح بعض الأخطاء ضمن القيم؛

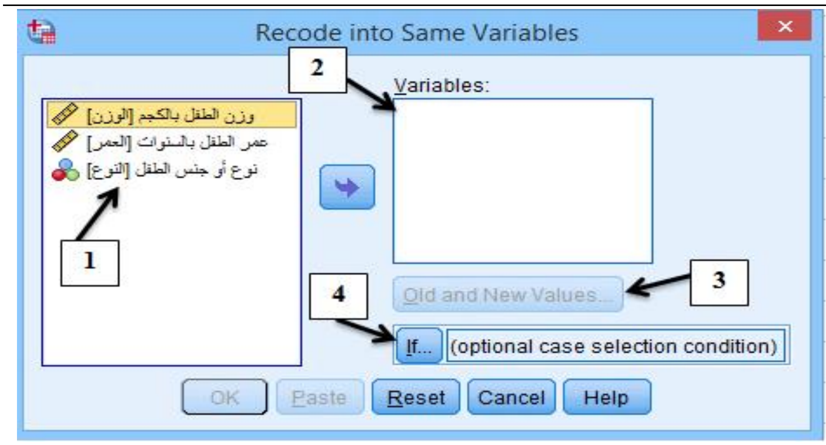
فمثلا عند اكتشاف وجود بعض الأخطاء في إدخال قيم أحد المتغيرات التي تمثل الحالة الاجتماعية بحيث أن بعض الأشخاص قد تم تسجيل مستواهم الدراسي بأنه "ماجستير" وهو في الحقيقة "دكتوراه"، فإن هذا يستدعي تصحيح هذا الخطأ يدويا أي تعديل القيم الخاصة بهؤلاء الأشخاص فقط، أما إذا كان المطلوب مثلا هو تغيير قيم المتغير الخاص بالمستوى الدراسي بحيث تتغير كل القيم المناظرة للقيمة "ماجستير" والقيمة "دكتوراه" إلى القيمة "فوق الجامعي"، فهنا يمكن استخدام خيار إعادة الترميز التلقائي.

ويمكن في برنامج SPSS تنفيذ إعادة الترميز لنفس المتغير، (أي تغيير قيم المتغير الأصلي نفسه)، أو تغيير القيم وتعريفها باسم متغير جديد. وسنبدأ بتطبيق الخيار الأول وهو تغيير قيم المتغير نفسه.

1.5.2 إعادة ترميز القيم لنفس المتغير (Recode into Same Variable)

في هذا الخيار، سيتم تغيير قيم المتغير نفسه بحيث أنه سيتم استبدال القيم الأصلية بالقيم الجديدة، لذلك لا يتم استخدام هذا الخيار إلا عند التأكد من عدم احتياجنا للقيم الأصلية للمتغير لاحقا. وكمثال تطبيقي، لنأخذ ملف البيانات المحفوظ لدينا باسم "مثال1".

لنفرض أننا نرغب بإعادة ترميز الذكور بالقيمة 0 بدلا من القيمة الأصلية 1، وإعادة ترميز الإناث بالقيمة 1 بدلا من القيمة الأصلية 2، فنقوم بالآتي؛ قم بفتح الملف "مثال1"، ثم قم من شريط الأوامر العلوي باختيار أمر تحويل (Transform) ثم اختر **إعادة ترميز في نفس المتغيرات** (Recode into Same Variables).



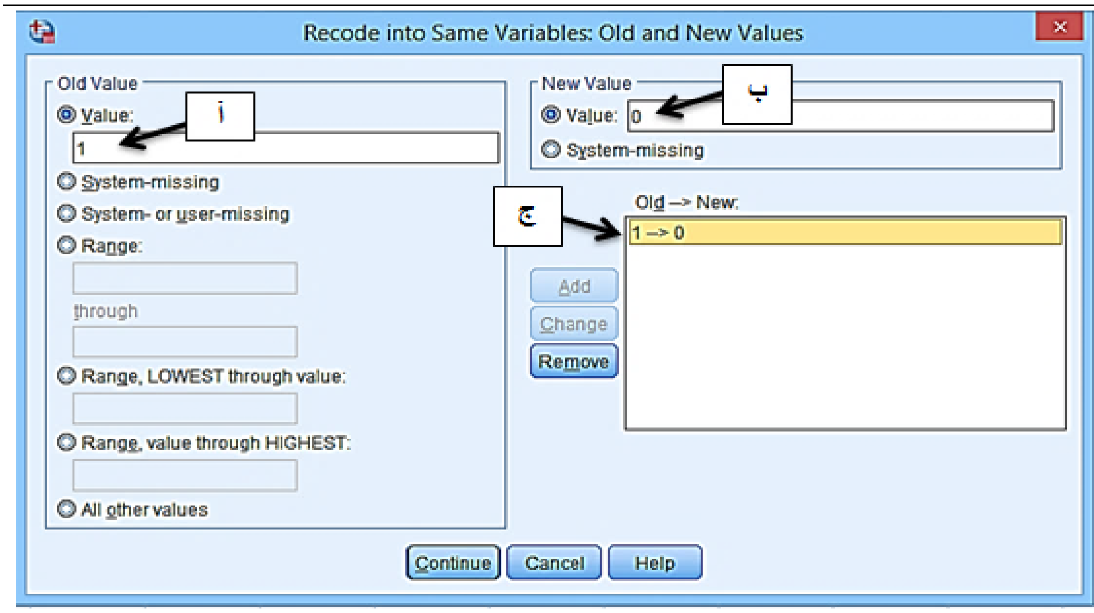
شكل 14.2: النافذة الخاصة بإعادة الترميز في نفس المتغير للبيانات "مثال1".

ستظهر عندها نافذة جديدة، كما يُشاهد في الشكل (14.2)، قم في هذه النافذة باختيار المتغير المطلوب إعادة ترميزه، وهو متغير "النوع" في مثالنا، واضغط سهم النقل لنقله إلى خانة المتغيرات المشار إليها بالرقم (2). بعد ذلك اضغط على خيار تعريف القيم القديمة والجديدة (Old and New Values) والمشار إليه بالرقم (3)، فتظهر نافذة جديدة، (الشكل (15.2))، والتي سيتم فيها إعادة ترميز القيم.

ملاحظة:

بالنسبة للخيار (4)، في الشكل (14.2)، وهو الدالة الشرطية إذا (If)، فقد سبقت الإشارة إليه حيث يمكن استخدامه في وضع قيد أو شرط على عملية تغيير القيم ضمن المتغيرات بحيث يتم اعتبار أو استثناء قيم محددة بناء على الشرط أو الشروط الموضوعة من قِبل المستخدم.

حيث أن التغيير المطلوب في متغير "النوع" هو تغيير القيمة (القديمة) 1 إلى (الجديدة) 0 والقيمة (القديمة) 2 إلى (الجديدة) 1، فنقوم بإدخال القيمة 1 في خانة القيمة القديمة (Old Value) المشار إليها بـ (أ)، وإدخال القيمة 0 في خانة القيمة الجديدة (New Value) المشار إليها بـ (ب) ثم نضغط إضافة (Add) فيتم إعادة الترميز كما يُشاهد في الخانة (ج). وبنفس الكيفية، يتم إدخال التغيير الثاني للمتغير؛ وهو القيمة 2 في (أ) والقيمة 1 في (ب) فيظهر الترميز الجديد في الخانة (ج) أسفل الترميز السابق.[1]



شكل 15.2: نافذة تعريف القيم القديمة والجديدة للمتغير في خيار إعادة الترميز في نفس المتغير.

قم بعد ذلك بالضغط على استمرار (Continue) للعودة للنافذة السابقة (الشكل (14.2))، وفيها اضغط موافق (OK) للانتهاء. وبعدها ستلاحظ تنفذ التغير المطلوب في قيم متغير "النوع". إلا أنه في هذا المثال، يجب تعديل

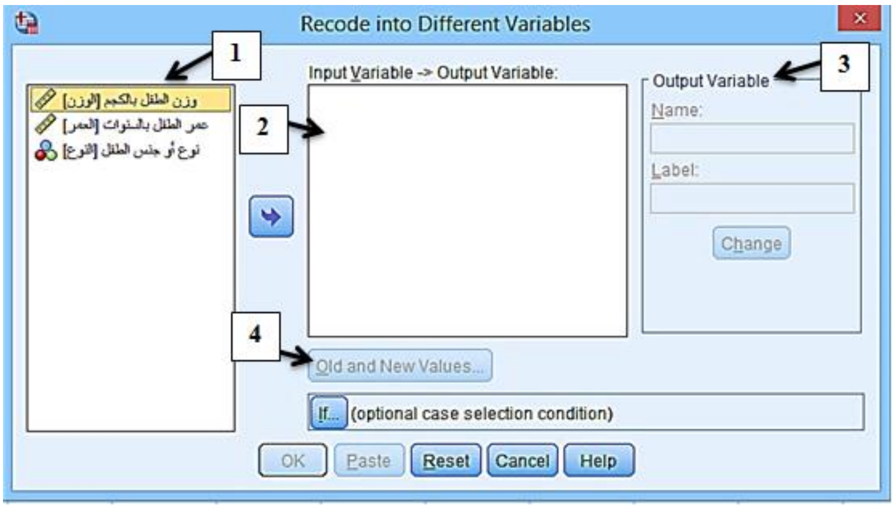
[1] يمكن بهذه الطريقة إعادة ترميز أكثر من قيمتين لأي متغير، وأما الخيارات الأخرى التي تظهر في النافذة، فيكون لها نفس الدور كما هو موضح في الشكل السابق (11.2).

تصنيف قيم المتغير (Values) في اسلوب عرض المتغيرات بحيث تناظر الرموز الجديدة 0 و1 التصنيف ذكر وأنثى على الترتيب. لذلك قم بتعيين القيمة 0 للتصنيف "ذكر" والقيمة 1 للتصنيف "أنثى" وقم بعدها بإلغاء (Remove) التصنيفات القديمة، ثم اضغط موافق (OK). وبهذا نكون قد نفذنا عملية إعادة ترميز القيم في نفس متغير.

2.5.2 إعادة ترميز القيم في متغير مختلف (Recode into Different Variable)

في هذه الحالة سيتم تعريف متغير جديد قيمه تمثل الترميز المطلوب الذي تم باستخدام المتغير الأصلي، بمعنى أنه سيتم تعريف متغير أو متغيرات إضافية في ملف البيانات. ويمكن استخدام ملف البيانات "مثال2"، كمثال توضيحي لعملية إعادة الترميز.

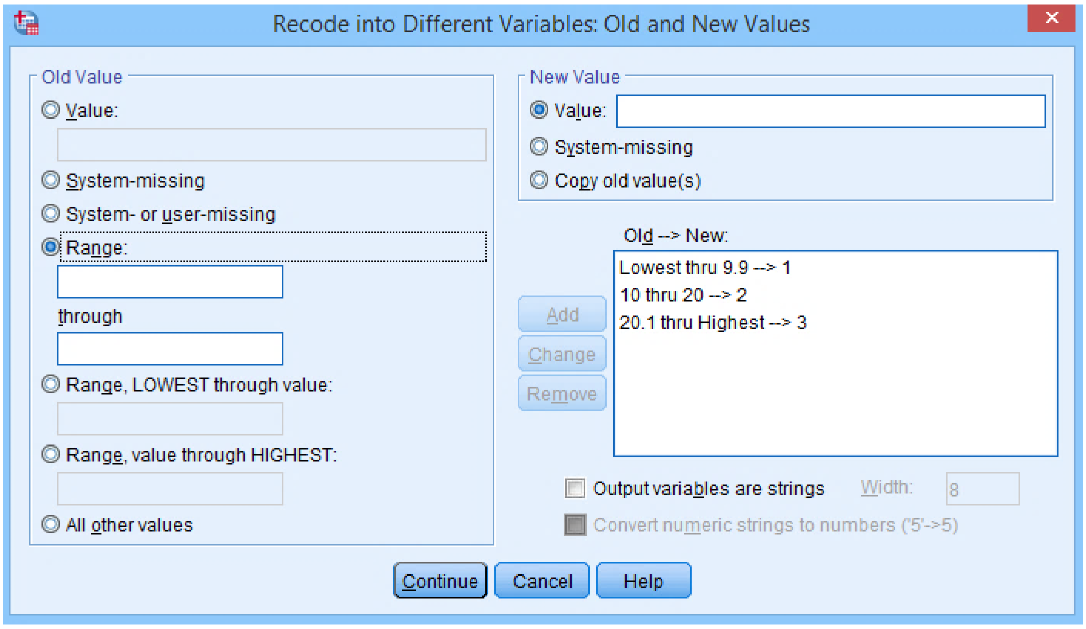
لنفرض أننا نرغب بتعريف متغير جديد يمثل تصـــنيف لأوزان الأطفال في البيانات "مثال2"، بحيث يتم إعطاء الرمز أو القيمة 1 للأطفال الذين تقل أوزانهم عن 10 كجم، والقيمة 2 للأطفال الذين تتراوح أوزانهم ما بين 10 و20 كجم، والقيمة 3 للأطفال الذين تزيد أوزانهم عن 20 كجم. لتنفيذ ذلك، قم أولا بفتح ملف البيانات "مثال2"، ثم اختيار أمر تحويل (Transform) من شـــريط الأوامر العلوي، ثم اختر **إعادة ترميز في متغير مختلف** (Recode into Different Variables). ستظهر عندها نافذة إعادة الترميز في متغير جديد، كما يُشاهد في الشكل (16.2).



شكل 16.2: النافذة الخاصة بإعادة الترميز في متغير مختلف للبيانات "مثال2".

سنقوم باختيار المتغير المطلوب، وهو متغير "الوزن" من الخانة (1)، ثم ننقله إلى الخانة (2) باستخدام سهم النقل. وفي الخانة (3)، سيتم اختيار الاسم المرغوب للمتغير الجديد ووصفه، ولنقم بتسمية المتغير الجديد بالاسم "الوزن2" وكتابة "تصنيف الوزن" في مربع الوصف (Label)، ثم نضغط على أمر **تغيير** (Change)، فنلاحظ عندها حدوث هذا التغيير في الخانة (2) حيث سيظهر فيها النص "الوزن --> الوزن2".

بعد ذلك سنقوم بتعريف القيم التي نود إعادة كتابتها في المتغير الجديد باستخدام قيم متغير "الوزن"، فنقوم باختيار أمر تعريف القيم القديمة والجديدة (Old and New Values) والمشار إليه بالرقم (4) في الشكل (16.2)، فتظهر نافذة جديدة كما يُرى في الشكل (17.2).



شكل 17.2: نافذة تعريف القيم القديمة والجديدة للمتغير في خيار إعادة الترميز في متغير مختلف.

في هذه النافذة، (شكل (17.2))، سنقوم بتعيين القيم 1، 2، و 3 للتصنيفات الثلاثة؛ الأطفال الذين تقل أوزانهم عن 10 كجم، الأطفال الذين تتراوح أوزانهم ما بين 10 و 20 كجم، والأطفال الذين تزيد أوزانهم عن 20 كجم على الترتيب.

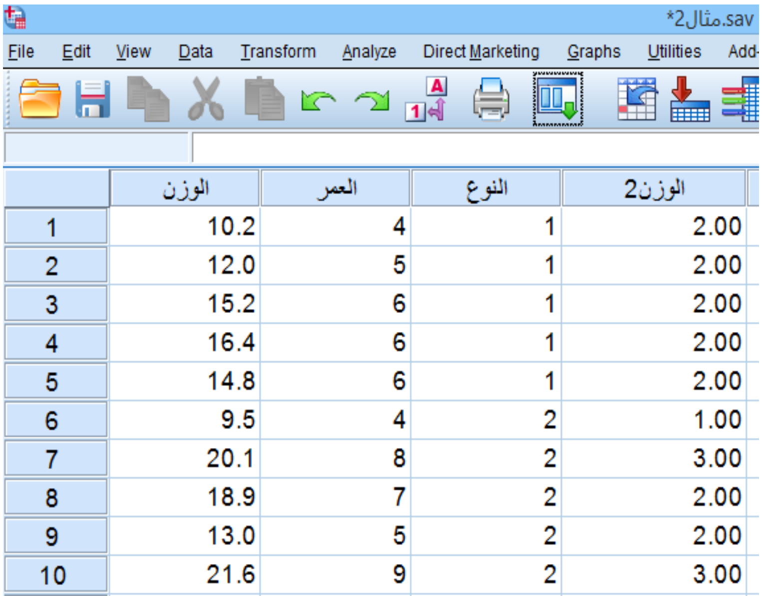
لتعيين القيمة الأولى، والتي تشمل كل الأوزان التي هي أقل من 10 كجم، نحتاج لاستخدام الخيار **المدى، القيمة فأقل** (Range, LOWEST through value). فنكتب أولا القيمة المحددة للتصنيف الأول وهي القيمة 1 في خانة القيمة الجديدة (New Value)، ثم نكتب القيمة 9.9، (وهي القيمة[1] الأقل من 10 كجم مباشرة في متغير الوزن)، ثم نضغط إضافة (Add). وبهذا نكون قد قمنا بتعيين القيمة الأولى وهي 1 إلى التصنيف الأول وهو الأوزان الأقل من 10 كجم. وبنفس الطريقة سيتم تعيين القيم التالية للتصنيفات التالية.

لتعيين القيمة الثانية، وهي القيمة 2 للتصنيف الثاني وهو الأوزان من 10 كجم إلى 20 كجم، نكتب القيمة 2 في خانة القيمة الجديدة (New Value) ثم نضغط على الخيار المدى (Range) فتظهر خانتان فارغتان أسفل منه. يتم إدخال القيمة 10 في الخانة العليا والقيمة 20 في الخانة السفلى والضغط على (Add).

[1] ننوه هنا إلى أنه إذا ما تم كتابة القيمة 10 في خانة (Range, LOWEST through value) فإن هذا التصنيف سيشمل على الوزن 10 كجم ضمنه.

ولتعيين القيمة الثالثة، وهي القيمة 3 للتصنيف الثالث، وهم الأطفال الذين تزيد أوزانهم عن 20 كجم، نكتب القيمة 3 في خانة القيمة الجديدة (New Value) ثم نضغط على الخيار **المدى، القيمة فأكبر من** (Range, value through HIGHEST) ونكتب فيه القيمة 20.1، (لأن القيمة 20 كجم هي ضمن نطاق التصنيف الثاني)، ثم نضغط على إضافة. وبعد الانتهاء من تعيين كل القيم المحددة للتصنيفات السابقة نضغط على استمرار (Continue).

ستتم العودة بعد ذلك للنافذة السابقة، (شكل (16.2))، نقوم بعدها بالضغط على (OK) فيظهر المتغير الجديد "الوزن2" في نافذة عرض البيانات في الملف "مثال2" كما هو موضح في الشكل (18.2).



	الوزن	العمر	النوع	الوزن2
1	10.2	4	1	2.00
2	12.0	5	1	2.00
3	15.2	6	1	2.00
4	16.4	6	1	2.00
5	14.8	6	1	2.00
6	9.5	4	2	1.00
7	20.1	8	2	3.00
8	18.9	7	2	2.00
9	13.0	5	2	2.00
10	21.6	9	2	3.00

شكل 18.2: عرض بيانات ملف "مثال2" بعد تعريف المتغير "الوزن2".

3.5.2 إعادة الترميز التلقائي (Automatic Recoding)

في هذا النوع من إعادة الترميز ، يتم تلقائيا تعيين قيم افتراضية مناظرة لقيم المتغير الأصلية بحيث يأخذ كل تصنيف قيمة تسلسلية هي (1، 2، 3، ...). ويتم عادة استخدام هذا الأسلوب من إعادة الترميز عند الحاجة لتغيير قيم مجموعة كبيرة من المتغيرات دفعة واحدة أو الحاجة لتعريف متغير جديد يكون نسخة رقمية من المتغير الأصلي أو لغير ذلك من الأسباب. ولاستخدام هذا الأسلوب، لنأخذ في الاعتبار المثال التالي، (الجدول (3.2))، والذي يتكون من المتغيرين "التقدير" و"القسم" واللذان يمثلان التقدير الدراسي للطالب والقسم العلمي الذي ينتمي له لعينة من 15 طالب. قم بإدخال البيانات في برنامج SPSS في ملف جديد[1] باسم "مثال4". وسنقوم بإعادة ترميز كلا من المتغيرين؛ "التقدير" و"القسم" باستخدام إعادة الترميز التلقائي.

[1] عند إدخال أسماء الأقسام للمتغير "القسم" يجب زيادة عدد الخانات في عامود العرض (Width) لتستوعب أسماء الأقسام الطويلة، (12 كعدد خانات سيكون مناسبا).

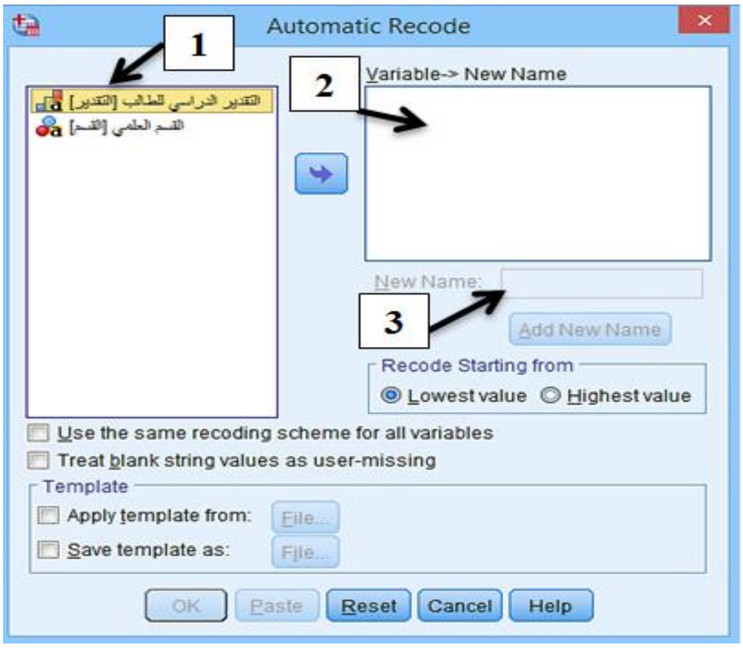
جدول 3.2: البيانات الخاصة بملف البيانات "مثال4".

التقدير	B	A	B	B	C	D	A	D	C	C	C	B	A	C	B
القسم	الإحصاء	الرياضيات	الإحصاء	الفيزياء	الكيمياء	علم النبات	الإحصاء	علم	علم النبات	الرياضيات	الكيمياء	الإحصاء	الإحصاء	الرياضيات	الكيمياء

في شريط الأوامر العلوي، قم باختيار أمر تحويل (Transform) ثم اختر **إعادة الترميز التلقائي** (Automatic Recode) فتظهر النافذة الخاصة بهذا الأمر كما هو موضح في الشكل (19.2) أدناه. في الخانة المشار لها بالرقم (1) سنقوم باختيار المتغير الأول الذي سيتم إعادة ترميزه تلقائيا، وهو متغير "التقدير"، ثم نقوم بنقله إلى الخانة (2) باستخدام سهم النقل. بعد ذلك نقوم في خانة الاسم الجديد (New Name) الخانة (3) بكتابة الاسم المرغوب للمتغير الجديد (الذي سيحصل على الترميز التلقائي)، وليكن "التقدير2" ثم نضغط على أمر إضافة اسم جديد (Add New Name) فتظهر لنا في الخانة (2) العبارة؛ "التقدير --> التقدير2".

بنفس الكيفية يتم اختيار المتغير الثاني "القسم" من الخانة (1) ونقله إلى الخانة (2) وتسميته، على سبيل المثال، "القسم2" والضغط على أمر (Add New Name) فتظهر العبارة؛ "القسم --> القسم2" في الخانة (2).

وننوه هنا إلى إمكانية اختيار الترتيب المطلوب لإعادة الترميز تصاعديا أو تنازليا باستخدام خيار **إعادة الترميز بداية من** (Recode Starting from)، حيث يستخدم خيار القيمة الأقل (Lowest value) للترتيب التصاعدي، وخيار القيمة الأكبر (Highest value) للترتيب التنازلي.



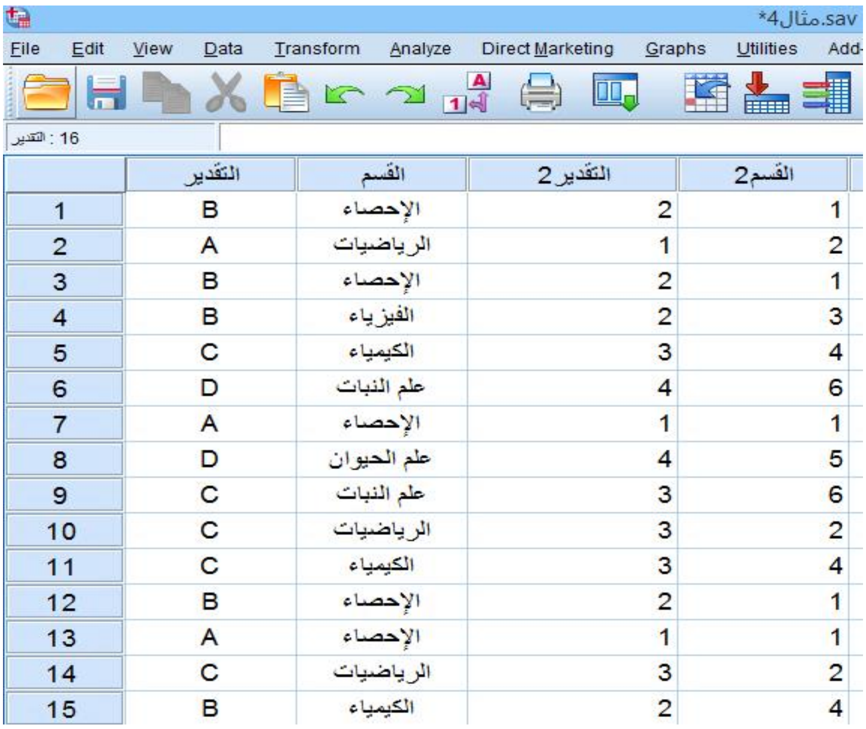
شكل 19.2: النافذة الخاصة بإعادة الترميز التلقائي لملف البيانات "مثال4".

ملاحظة:

في نافذة إعادة الترميز التلقائي، (الشكل (19.2))، يمكن استخدام الخيار الخاص بحفظ وإعادة تطبيق نمط إعادة الترميز أي **القالب** (Template) الموجود في أسفل النافذة عند التعامل مع بيانات أخرى مستقبلا.

بعد الانتهاء من اختيار المتغيرات وتعريف أسماء المتغيرات الجديدة نضغط موافق (OK) فيتم التنفيذ ويظهر المتغيران الجديدان "التقدير2" و"القسم2" في ملف البيانات "مثال4"، (الشكل (20.2)). ويمكنك رؤية نمط الترميز للمتغيرين من خلال نافذة المخرجات (Output).

ولاحظ أن ترميز المتغير "التقدير" قد تم بحيث أن قيم المتغير (A, B, C, D) تناظر القيم (1, 2, 3, 4) في المتغير الجديد "التقدير2" على الترتيب. وأما في المتغير "القسم2"، فقد تم تعيين القيم (من 1 إلى 6) بحسب عدد الأقسام أبجديا، بداية من قسم الإحصاء ممثلا بالقيمة 1 وانتهاء بقسم علم النبات ممثلا بالقيمة 6.



	التقدير	القسم	التقدير2	القسم2
1	B	الإحصاء	2	1
2	A	الرياضيات	1	2
3	B	الإحصاء	2	1
4	B	الفيزياء	2	3
5	C	الكيمياء	3	4
6	D	علم النبات	4	6
7	A	الإحصاء	1	1
8	D	علم الحيوان	4	5
9	C	علم النبات	3	6
10	C	الرياضيات	3	2
11	C	الكيمياء	3	4
12	B	الإحصاء	2	1
13	A	الإحصاء	1	1
14	C	الرياضيات	3	2
15	B	الكيمياء	2	4

شكل 20.2: عرض بيانات ملف "مثال4" بعد تعريف المتغيران "التقدير2" و"القسم2".

4.5.2 تعريف دالة جديدة لمتغير (Compute Variable)

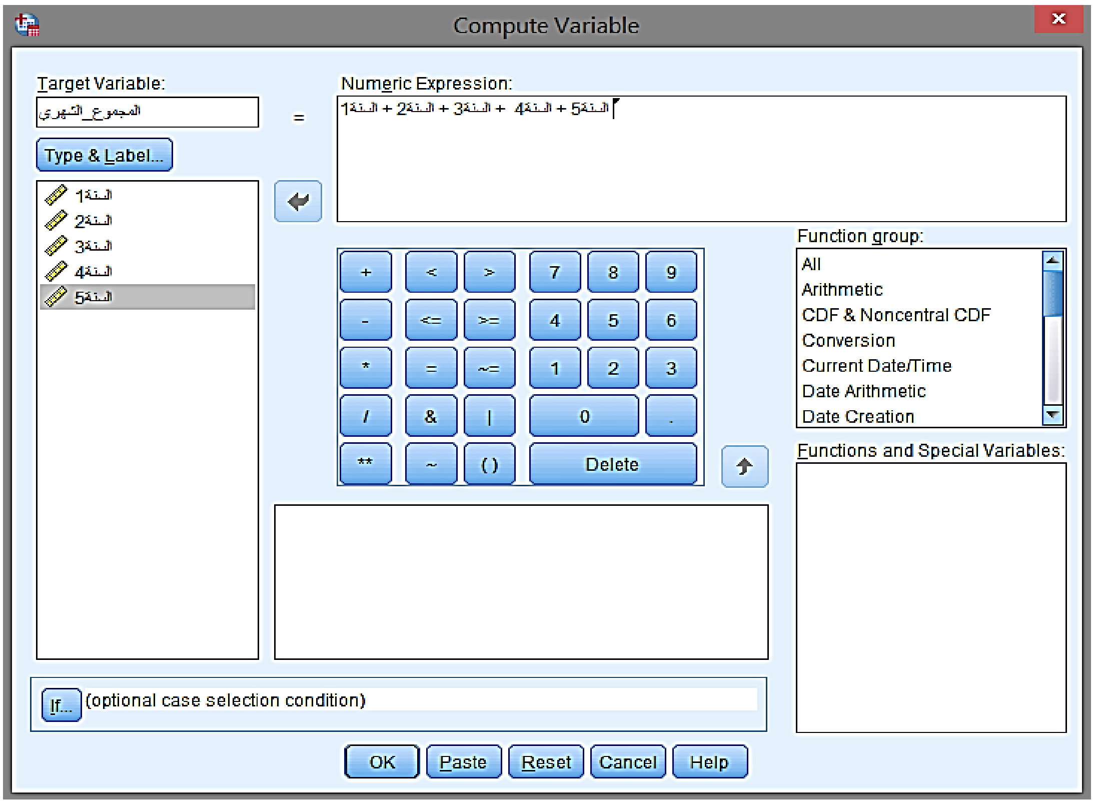
يُعد هذا الأمر في برنامج SPSS بمثابة الحالة العامة للترميز أو استخدام دوال حسابية وإحصائية في تعريف متغير جديد اعتمادا على المتغير الأصلي. واستخدام هذا الأمر يتضمن مجالات واسعة تتدرج من تعريف الدوال البسيطة، (مثل العمليات الحسابية البسيطة؛ جمع، طرح، ضرب، قسمة)، إلى تعريف الدوال الرياضية والإحصائية الأكثر تعقيدا.

وسنقوم هنا بتقديم بعض الأمثلة على استخدام هذا الأمر من خلال البيانات في الجدول التالي، (جدول (4.2))، والذي يتضمن أسعار النفط العالمية الشهرية لمزيج برنت بالدولار الأمريكي لخمسة سنوات افتراضية. وسنقوم بإدخال هذه البيانات في ملف بيانات جديد في برنامج SPSS باسم "أسعار النفط".

جدول 4.2: أسعار النفط العالمية الشهرية لمزيج برنت بالدولار الأمريكي لخمسة سنوات.

	السنة1	السنة2	السنة3	السنة4	السنة5
1	42	45	48	58	61
2	41	46	50	58	61
3	42	45	51	56	62
4	43	47	52	57	64
5	42	47	52	58	64
6	44	48	52	57	65
7	43	47	54	59	63
8	42	49	53	60	63
9	40	47	55	60	62
10	44	48	54	58	63
11	45	46	55	60	64
12	44	49	57	59	65

في شريط الأوامر العلوي، قم باختيار Transform>Compute Variable فتظهر نافذة حساب المتغير كما يظهر في الشكل (21.2).



شكل 21.2: نافذة تعريف دالة جديدة لمتغير، (حساب متغير)، لملف بيانات "أسعار النفط".

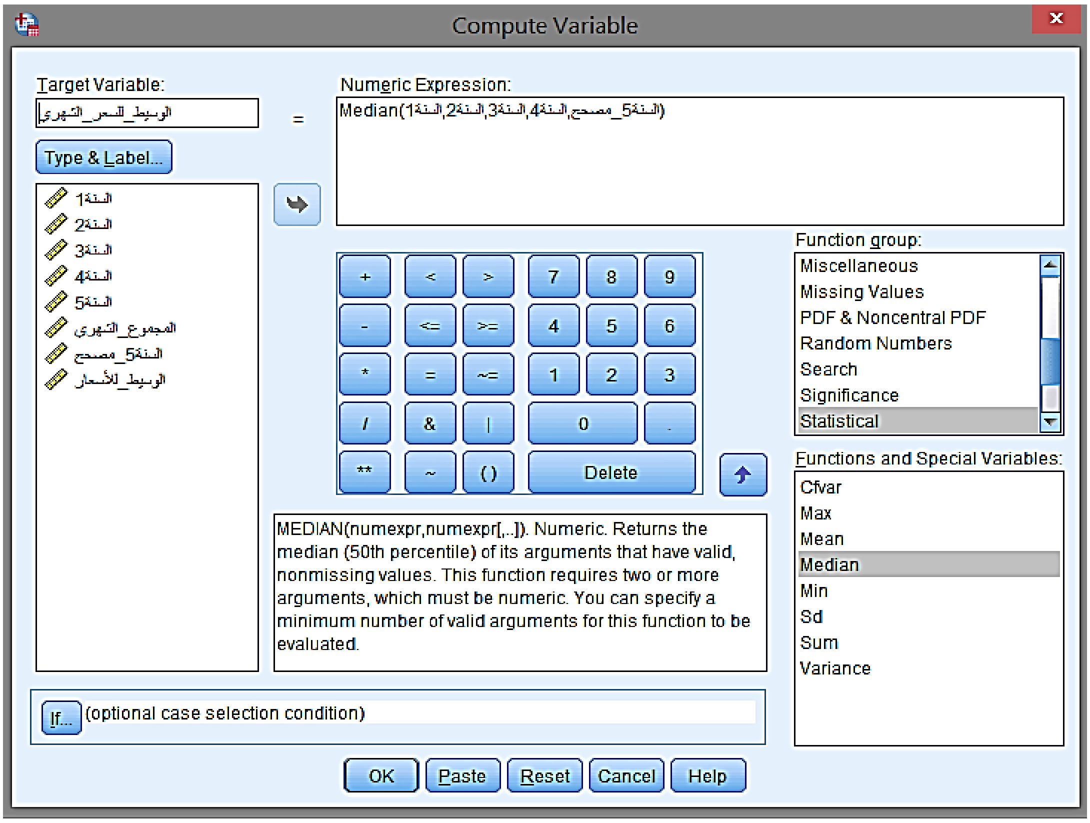
لنفرض الآن أننا نريد تعريف متغير جديد يمثل المجموع الشهري لأسعار النفط خلال السنوات الخمس، فنقوم في نافذة حساب المتغير بكتابة اسم المتغير الجديد، وليكن "المجموع_الشهري"، في مربع المتغير المستهدف (Target Variable)، ثم ندخل الدالة أو العملية الحسابية المطلوبة في مربع الدالة الحسابية (Numeric Expression) وهي مجموع القيم لكل شهر (والذي يمثله صف المشاهدات) في المتغيرات الخمسة، وهكذا فإن العملية المطلوبة سيتم كتابتها كمجموع للمتغيرات الخمسة كما هو ظاهر في الشكل (21.2).

ويتم تنفيذ ذلك عن طريق اختيار المتغير الأول (السنة1) من مربع المتغيرات في يسار النافذة ونقله بالسهم إلى مربع الدالة الحسابية ثم الضغط على إشارة الجمع (+) في لوحة المفاتيح في وسط النافذة، ثم اختيار المتغير الثاني (السنة2) ونقله إلى مربع الدالة الحسابية ثم الضغط على إشارة الجمع وهكذا حتى اختيار المتغير الخامس (السنة5). بعد ذلك نضغط موافق (OK) للتنفيذ، وفي ملف البيانات الحالي، "أسعار النفط"، سيظهر المتغير الجديد باسم "المجموع_الشهري".

كمثال آخر، لنفرض أن الأسعار في السنة الخامسة كان بها خطأ في الرصد حيث أنها كانت أعلى من القيم الصحيحة بمقدار دولارين، ونريد الآن تصحيح هذا الخطأ عن طريق تعريف متغير جديد. لذلك نقوم في نافذة حساب المتغير بكتابة اسم المتغير الجديد، وليكن "السنة5_مصحح"، في مربع المتغير المستهدف ثم نقوم باختيار المتغير الذي سيتم التعامل معه، وهو "السنة5" من مربع المتغيرات وننقله إلى مربع الدالة الحسابية ثم نختار إشارة الطرح (−) في لوحة المفاتيح في وسط النافذة ونكتب القيمة 2 بعدها، ثم نضغط موافق فنحصل على المتغير الجديد المطلوب.

ويمكن استخدام بعض الدوال الإحصائية في تعريف متغير جديد، فمثلا إذا ما أردنا حساب متغير جديد يمثل قيمة الوسيط (Median) لأسعار النفط الشهرية، (أي حساب الوسيط لأسعار النفط لكل شهر من أشهر السنة على مدى السنوات الخمسة)، فإننا نقوم بالآتي؛ في نافذة حساب المتغير بكتابة اسم المتغير الجديد، وليكن "الوسيط_للسعر_الشهري"، في مربع المتغير المستهدف ثم نقوم باختيار الدوال الإحصائية (Statistical) من مربع مجموعات الدوال (Function group) في يمين النافذة، فتظهر قائمة بالدوال الإحصائية المتوفرة في مربع الدوال والمتغيرات الخاصة (Functions and Special Variables) أسفل منها. قم باختيار دالة الوسيط (Median) من بين تلك الدوال عن طريق النقر المزدوج عليها، (أو سحبها بالفأرة إلى المربع المرغوب).

عندها ستظهر دالة الوسيط في مربع الدالة الحسابية مع علامتي استفهام، قم بإلغاء الأقواس وعلامتي الاستفهام ثم قم بفتح قوس جديد وانقل المتغيرات "السنة1"، "السنة2"، "السنة3"، "السنة4"، و"السنة5_مصحح" بعد القوس المفتوح مع مراعاة تغيير لغة لوحة المفاتيح في جهازك إلى اللغة الإنجليزية وإدخال فواصل بين المتغيرات، ثم إغلاق القوس بعد ذلك، كما يوضح الشكل (22.2) لنافذة حساب المتغير. بعد ذلك اضغط موافق للتنفيذ.



شكل 22.2: حساب الوسيط الشهري في نافذة تعريف دالة جديدة لمتغير، لملف بيانات "أسعار النفط".

وبالتالي سيحتوي ملف بيانات "أسعار النفط" على المتغيرات التي تظهر في الشكل (23.2).

	السنة1	السنة2	السنة3	السنة4	السنة5	المجموع_الشهري	السنة5_مصحح	الوسيط_للسعر_الشهري
1	42	45	48	58	61	254	59	48
2	41	46	50	58	61	256	59	50
3	42	45	51	56	62	256	60	51
4	43	47	52	57	64	263	62	52
5	42	47	52	58	64	263	62	52
6	44	48	52	57	65	266	63	52
7	43	47	54	59	63	266	61	54
8	42	49	53	60	63	267	61	53
9	40	47	55	60	62	264	60	55
10	44	48	54	58	63	267	61	54
11	45	46	55	60	64	270	62	55
12	44	49	57	59	65	274	63	57

شكل 23.2: ملف بيانات "أسعار النفط" بعد تعريف المتغيرات الجديدة.

وهكذا، وكما ذكرنا أعلاه، يمكن استخدام أمر تعريف دالة جديدة لمتغير (Compute Variable) بطرق عديدة يمكن أن تكون مفيدة جدا أثناء إجراء التحليل الإحصائي للبيانات.

5.5.2 الترتيب في مجموعات (Binning)

هذه الطريقة من إعادة الترميز للمتغيرات تشبه إعادة الترميز في متغير مختلف الذي تناولناه سابقا، إلا أنها تقوم "بقراءة" قيم مشاهدات المتغير وتعطي الخيار للمستخدِم بأن يحدد المجموعات أو الفترات التي يرغب بتصنيف المتغير الأصلي اعتمادا عليها يدويا أو تلقائيا.

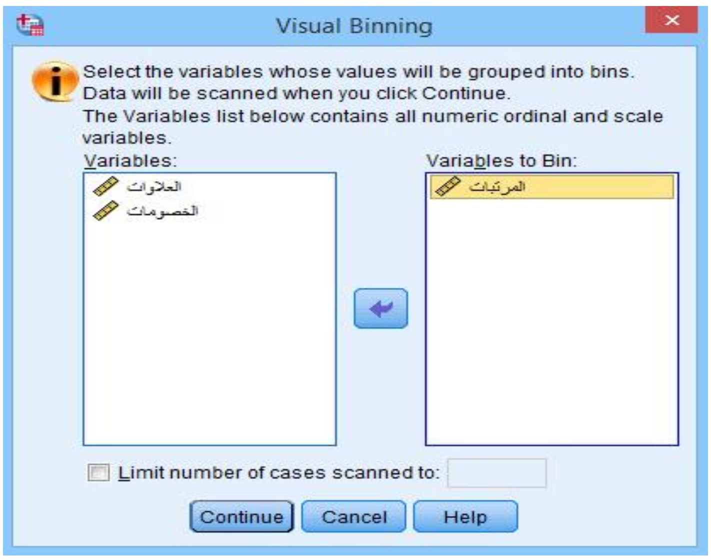
في الجدول التالي، (جدول (5.2))، لدينا بيانات خاصة بعدد من موظفي إحدى الشركات وهي عبارة عن ثلاثة متغيرات هي مرتب الموظف والعلاوة والخصم السنوي بالدينار الليبي. قم أولا بإدخال هذه البيانات في ملف في SPSS وحفظها باسم "مثال5".

جدول 5.2: البيانات الخاصة بمرتبات وعلاوات وخصومات الموظفين بالدينار الليبي. (الملف مثال5).

الموظف	المرتبات	العلاوات	الخصومات
1	480.500	120.050	25.000
2	635.700	165.500	70.750
3	390.550	95.050	15.500
4	345.590	85.000	12.750
5	730.600	170.750	75.750
6	980.050	185.500	90.000
7	1060.750	190.750	95.500
8	510.650	155.500	60.500
9	1150.500	195.750	100.500
10	560.400	160.750	65.000
11	490.900	110.050	20.500
12	870.750	180.000	85.750
13	910.950	190.500	95.000

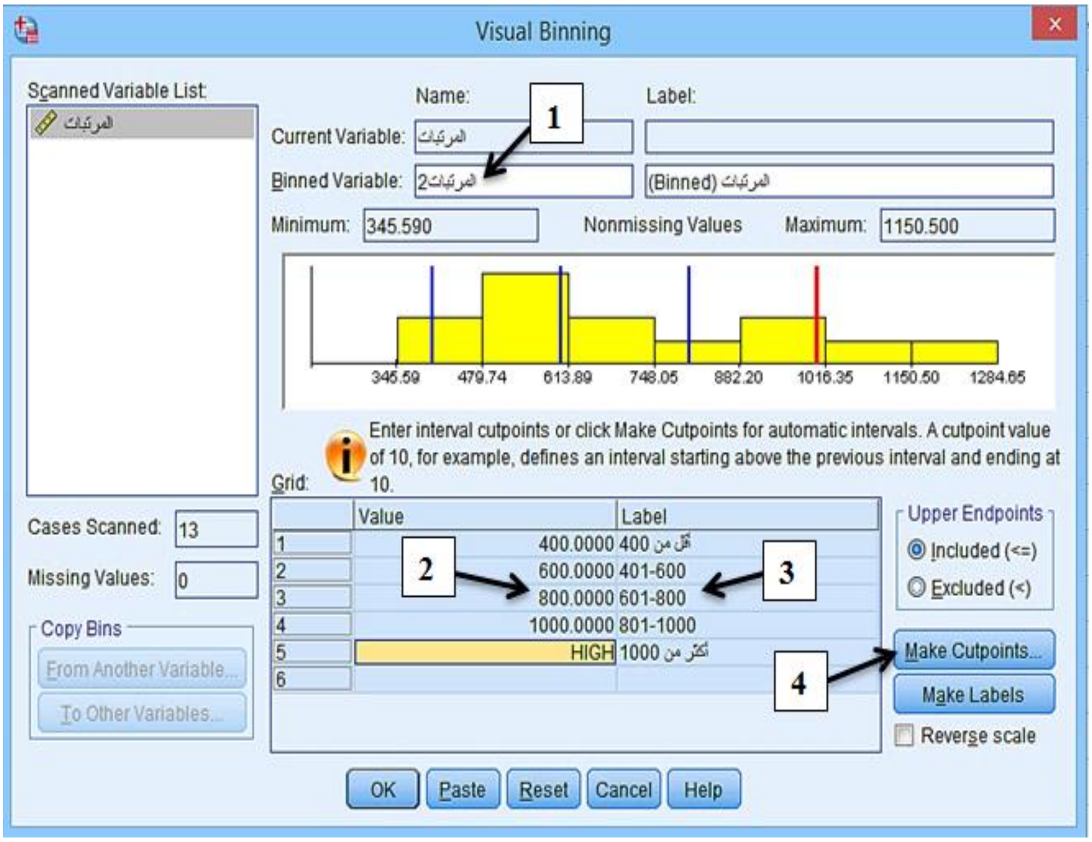
ولنفرض أننا نريد إعادة ترميز المتغير "المرتبات" إلى خمسة مجموعات أو فترات بحسب مقدار المرتب وهي (أقل من 400)، (401-600)، (601-800)، (801-1000)، و(أكثر من 1000) وذلك باستخدام طريقة الترتيب في مجموعات فنقوم بالآتي؛

في شريط الأوامر العلوي قم باختيار أمر تحويل (Transform)، ثم اختر **الترتيب في مجموعات (المرئي) (Visual Binning)**. ستظهر عندها النافذة الخاصة بتحديد المتغير المطلوب إعادة ترميزه، (كما يشاهد في شكل (24.2))، وهو في مثالنا "المرتبات" فنقوم باختياره والضغط على سهم النقل، ثم استمرار (Continue).



شكل 24.2: النافذة الخاصة بتحديد المتغير المطلوب إعادة ترميزه باستخدام طريقة الترتيب في مجموعات لملف البيانات "مثال5".

بعد ذلك ستظهر لنا نافذة أخرى، (كما يظهر في الشكل (25.2))، والتي سيتم من خلالها تعريف المجموعات المرغوب تكوينها من المتغير الأصلي.



شكل 25.2: نافذة تعريف المجموعات للمتغير "المرتبات" في الملف "مثال5".

وبعد ستظهر ستظهر نافذة صغيرة للتأكيد على أنه سيتم تكوين متغير واحد جديد، قم بالضغط على موافق (OK) فيظهر المتغير الجديد "المرتبات2" في ملف البيانات "مثال5" والذي يأخذ القيم من 1 إلى 5 والتي ترمز للمجموعة أو الفترة المناظرة. ولاحظ أنه إذا ما تم استخدام أيقونة عرض وصف المتغير  من شريط الأيقونات العلوي فسيظهر في عامود المتغير الجديد وصف المجموعات التي تم تسميتها (بداية من "أقل من 400" إلى "أكثر من 1000").

الفصل الثالث

التحليل الاستكشافي للبيانات في SPSS
(Exploratory Data Analysis (EDA) in SPSS)

في هذا الفصل، سيتم تناول المقاييس والطرق الإحصائية المختلفة[1] وتوضيح كيفية استخدامها في برنامج SPSS. هذه المقاييس، التي تندرج تحت مسمى **الإحصاء الوصفي** (Descriptive Statistics) مصحوبة ببعض التقديرات الاستدلالية البسيطة تُعرف حديثا **بالتحليل الاستكشافي للبيانات** (Exploratory Data Analysis (EDA))، وتُعد المدخل الرئيسي لبوابة علم الإحصاء التطبيقي، حيث أنها توفر للباحث أو المستخدِم للأساليب الإحصائية المعلومات الأولية التي تمكنه من استكشاف ما تمثله أو تصفه البيانات. وللوصول للاستخدام الأمثل لهذه المقاييس الإحصائية، لابد للقارئ أولا من التعرف على أنواع البيانات الأساسية المستخدمة في التحليل الإحصائي.

1.3 أنواع البيانات (Data Types)

تمثل **البيانات** (Data) المادة الخام التي تزودنا بالمعلومات حول المجتمع الذي جمعت منه بعد استخدام الأسلوب أو الأساليب الإحصائية المناسبة. ونقصد بالمجتمع هنا المصدر الذي تم جمع[2] البيانات منه. لهذا، فإن مرحلة التعرف على البيانات في مصدرها تعد من المراحل الهامة في التحليل الإحصائي الصحيح.

إن لفظ "بيانات" كثيرا ما يظهر في الكتب والمنشورات ووسائل الإعلام والتعاملات اليومية في العموم، وهو مصطلح مرن شامل، فالبيانات الشخصية التي تضم الاسم، العمر، النوع، الحالة الاجتماعية، المستوى التعليمي، وغيرها هي نوع من البيانات، والتاريخ الطبي للمريض والذي يضم القراءات الخاصة به مثل الوزن، نسبة السكر، ضغط الدم، درجات الحرارة المسجلة، وغيرها هي بيانات، وكذلك المشاهدات الناتجة عن تجربة كيميائية لقياس التفاعل الناتج عن دمج بعض المحاليل هي أيضا بيانات. ولذلك فإنه يمكن القول بأن أي ظاهرة أو دراسة أو تجربة أو مراقبة لسير عملية معينة يمكن أن ينتج عنها جميعا بيانات، فالبيانات هي المقياس الفعلي الذي يصف أو يُعبّر عما حدث في المجتمع أو الظاهرة.

---

[1] باعتبار أن القارئ مُلم بالحد الأدنى للمفاهيم والطرق الإحصائية الأساسية لكل من الإحصاء الوصفي (Descriptive Statistics)، والإحصاء الاستدلالي (Inferential Statistics).

[2] نشير هنا إلى أن معظم النظريات الإحصائية تتعامل مع البيانات على أنها **عينة** (Sample) أو عينات قد سُحبت من **مجتمع** (Population) أو مجتمعات محددة باستخدام أساليب متعددة من **المعاينة العشوائية** (Random Sampling).

وفي كثير من الأحيان، قد يخلط البعض بين مفهوم "البيانات" ومفهوم "المعلومات"، وهما في الواقع مختلفان على الأقل من وجهة النظر الإحصائية، فعلم الإحصاء (في أبسط صُورِه) يتعامل مع البيانات كما أسلفنا على أنها المادة الخام التي ستصف و/أو تحلل الظاهرة المدروسة، كما نرى في المثال التوضيحي الذي سنعرضه في الخطوات التسلسلية التالية:

1. لدينا بيانات؛ (درجات مجموعة من الطلبة (من 100 درجة) في مقرر اللغة العربية: 33، 42، 59، 61، ...، 25).
2. نستخدم الإحصاء الاستكشافي (مثلا: معدل الطلبة أو الوسيط يساوي 54 درجة).
3. حصلنا على معلومات؛ (مستوى الطلبة في مقرر اللغة العربية يعتبر ضعيفا).
4. يمكن طرح التساؤل: ماهي الأسباب؟
5. يمكن جمع المزيد من البيانات واستخدام الإحصاء التحليلي الاستدلالي للحصول على معلومات إضافية وتفسير ظاهرة ضعف الطلبة في هذا المقرر.

والبيانات الإحصائية، عادة ما تأخذ شكلا محددا يتألف من جدول (مكون من صفوف وأعمدة) بحيث تكون **المشاهدات** (Observations) هي مكونات الصفوف، و**المتغيرات** (Variables) هي مكونات الأعمدة. والمتغير في جدول البيانات يشير إلى الصفة المميزة لشيء يمكن التعبير عنه بصورة كمية أو وصفية، فأي تجربة أو ظاهرة يمكن قياسها بوحدة قياس مُعرّفة أو تصنيفها (إلى مستويات محددة مسبقا) يمكن أن يعبر عنها بمتغير؛ فالطول (بالسنتيمتر أو القدم) لمجموعة من الأشخاص هو متغير، والعمر الاستهلاكي (بالأشهر أو السنوات) لجهاز كهربائي هو متغير، وتقديرات الطلبة (ممتاز، جيد جدا، جيد، مقبول، راسب) تمثل أيضا متغير.

وباستخدام الأساليب الإحصائية المناسبة، يتم التعامل مع هذه البيانات بصورة استكشافية أو تحليلية (استدلالية) بغية الحصول على المعلومات المطلوبة بلغة الأرقام (المقاييس) أو بالرسومات التوضيحية. وطرق التعامل الإحصائي مع البيانات واستخلاص المعلومات المباشرة وغير المباشرة منها يعتمد على طبيعة ونوع تلك البيانات، وينبغي على الباحث مراعاة تلك الطبيعة جيدا بهدف تحديد الأسلوب الإحصائي المناسب للتعامل معها. ويمكن من حيث الطبيعة أن نقسم أنواع البيانات (أي المتغيرات) إلى نوعين رئيسيين هما؛ **البيانات الكمية**[1] (Quantitative Data) **والبيانات النوعية** أو **الوصفية** (Qualitative Data)، والتي قد تسمى أيضا **بالبيانات القطاعية** أو **التصنيفية** (Categorical Data).

[1] يُقصد بالبيانات أو المتغيرات الكمية تلك التي يمكن قياسها بوحدة قياس مُعرّفة، (مثل العمر، الوزن، الزمن، السعر، درجة الحرارة، نِسب النجاح، ...)، ويُفضّل استخدام مصطلح متغيرات كمية عن متغيرات رقمية لأن المتغيرات الوصفية التي لا يمكن قياسها بوحدات قياس، (مثل الجنس، اللون، آراء الناس، المشاعر الإنسانية، التقديرات الدراسية، ...)، يمكن التعبير عنها بقيم رقمية أيضا كما وضحنا في الفصل الأول، إلا أن ذلك لا يعني أنها أصبحت متغيرات كمية.

فالبيانات الكمية هي تلك البيانات التي تحتوي على مشاهدات تم قياسها ورصدها مباشرة بصورة قيم لها وحدة قياس كمية، مثل درجات الطلبة من 100 درجة، وأسعار النفط بالدولار، والمدة الزمنية المستغرقة لرحلة جوية بالساعات، وغيرها. وهذا النوع من البيانات تمثله متغيرات يمكن التعامل معها بواسطة العمليات الحسابية مباشرة.

أما البيانات التي يتم التعبير عنها بصفات أو تصنيفات، مثل النوع (ذكر، أنثى)، أو حالة الطقس المتوقعة (صحو، غائم، ممطر، ...)، أو رأي الجمهور في برنامج تلفزيوني معين (جيد، لا بأس به، سيئ)، فهذه البيانات تسمى بيانات نوعية أو قطاعية حيث أنها تأخذ قيما تحدد تصنيف المشاهدة إلى نوع أو قطاع محدد ولا يتم التعامل معها إحصائيا بالعمليات الحسابية المباشرة حتى وإن تم إعطاء هذه الصفات أو التصنيفات رموزا رقمية.

ويمكن تقسيم البيانات الكمية من حيث المقياس إلى نوعين أساسين هما:

1. **المقياس الفئوي** (Interval Scale): وهو الذي يسمح بترتيب المشاهدات على مقياس محدد بحيث يمكن تحديد القيمة الفعلية لكل مشاهدة، بدون اعتبار الصفر كقيمة. ومن الأمثلة على هذا النوع من البيانات قياس درجة الحرارة عدديا أو قياس بعض الفترات الزمنية.

2. **المقياس النسبي** (Ratio Scale): وهو مشابه للمقياس الفئوي إلا أنه يَعتبر الصفر من ضمن درجات القياس، لذلك فإنه في هذا المقياس يمكن حساب النسبة[1] بين قيمتين ضمن المفردات. وهذا النوع من المقاييس يكون شاملا للعديد من أنواع البيانات الكمية مثل السعر والوزن والطول والزمن وغيرها.

ملاحظة:

يمكن تقسيم المتغيرات الكمية من حيث طبيعة الأرقام إلى متغيرات **منفصلة** (Discrete)، ومتغيرات **متصلة** (Continuous). فالمتغيرات الكمية المنفصلة تأخذ قيما في مجموعات أعداد منتهية أو غير منتهية وقابلة للعد، مثل الأعداد الصحيحة (Integers). أما المتغيرات الكمية المتصلة فهي تأخذ قيما ضمن فترة من الأعداد، وتشمل مجموعة الأعداد الحقيقية (Real Numbers).

أما البيانات أو المتغيرات الوصفية فتنقسم إلى نوعين هما:

1. **المقياس الاسمي** (Nominal Scale): وهو يمثل تصنيف المشاهدات إلى مستويات أو قطاعات مختلفة، إلا أنه لا يمكن تحديد القيمة الفعلية لكل مستوى، ولا حتى ترتيب هذه المستويات بشكل تصاعدي أو تنازلي. فمثلا، لا يمكن القول أن الإجابة "نعم" هي أفضل من الإجابة "لا" فيما يتعلق بسؤال معين. ومن أمثلة هذا النوع من البيانات نوع الشخص (ذكر، أنثى)، وأسماء الأشياء والألوان وما شابه ذلك.

[1] كأن نقول مثلا أن طفلا عمره 10 سنوات هو أكبر بمرتين من طفل عمره 5 سنوات.

2. **المقياس الترتيبي أو الرُتبي** (Ordinal Scale): وهو على عكس النوع السابق، فطبيعته تسمح بترتيب المشاهدات وفق تدرج محدد من الأقل قيمة إلى الأكثر قيمة أو العكس، مثل تقديرات الطلبة (ممتاز، جيد جدا، جيد، ...)، إلا أنه لا يمكن حساب الفرق في القيمة بين أي مستويين من المستويات المرتبة.

وسواء كانت البيانات كمية أو وصفية أو مزيج بين الاثنين، فإنه يمكن التعامل معها من زاوية إحصائية بشكل مفرد، أي التعامل مع متغير واحد، وهو ما يُعرف **بالتحليل الأحادي** (Univariate Analysis)، أو التعامل مع عدة متغيرات في آن واحد، وهو ما يُعرف **بالتحليل المتعدد**[1] (Multivariate Analysis). والأساليب الإحصائية التي تُستخدم في التحليل الأحادي تعني أنه سيتم حساب تلك المقاييس لكل متغير على حدة، حتى وإن كانت مجموعة أو قاعدة البيانات متكونة من متغيرات عديدة.

ومثال على ذلك؛ بافتراض وجود بيانات بها ثلاثة متغيرات تمثل درجات طلبة في ثلاثة كليات مختلفة، فإنه إذا ما تم التعامل، (أي استخدام الأساليب الإحصائية)، مع درجات الطلبة لأي متغير بمفرده فإن ذلك يكون ضمن إطار التعامل مع التحليل الأحادي حتى وإن تم تنفيذ ذلك لكل المتغيرات. أما إذا ما تم استخدام أسلوب أو أساليب إحصائية تحتوي على صيغ تتعامل مع متغيرين[2] أو أكثر في نفس الوقت فإن ذلك يُسمى تحليل ثنائي، (في حالة استخدام متغيرين)، أو تحليل متعدد إذا ما تم إدراج أكثر من متغيرين في التحليل الإحصائي. وفيما يلي، سنقوم بإفراد مساحة لتوضيح كيفية تنفيذ تحليل البيانات الاستكشافي الأحادي والمتعدد باستخدام برنامج SPSS.

2.3 التحليل الاستكشافي الأحادي (Univariate EDA)

في معظم، إن لم يكن كل، الدراسات الإحصائية التي تتعلق بتحليل البيانات أو الدراسات الاجتماعية أو الاقتصادية أو الطبية أو غيرها، والتي تشتمل على جانب يُستخدم فيه التحليل الإحصائي للبيانات، يُفضل عادة استخدام ما يُعرف بالتحليل الاستكشافي للبيانات للتعرّف على **توزيع البيانات** (Data Distribution) وفحصها وذلك لعدة أسباب نذكر منها ما يلي:

1. كشف المشاكل والأخطاء التي قد توجد ضمن البيانات مثل **المشاهدات المتطرفة** (Outliers)، ابتعاد توزيع المتغيرات الكمية عن التوزيع الطبيعي، وجود أخطاء في ترميز البيانات، وجود **قيم مفقودة** (Missing Values)، وجود أخطاء في إدخال البيانات، وغيرها من المشاكل التي قد تتسبب في حصول الباحث على نتائج خاطئة أو مضللة.

[1] وتوجد حالة خاصة من التحليل المتعدد هو **التحليل الثنائي** (Bivariate Analysis).

[2] من أشهر الأساليب الإحصائية التي تتعامل مع متغيرين أو أكثر في آن واحد هي مقاييس التغاير والارتباط كما سنوضح لاحقا.

2. التأكد من مدى تبعية أو ملائمة البيانات للفرضيات التي قد يعتمد عليها استخدام معظم طرق التحليل الإحصائي الاستدلالي[1].
3. الحصول على معلومات أولية من البيانات تُعطي الباحث انطباعا عما تمثله الظاهرة محل الدراسة والنمط أو السلوك الذي تسلكه البيانات.
4. التعرّف على العلاقات البسيطة، إن وُجدت، بين متغيرات الدراسة وذلك لبناء النموذج أو النماذج الأفضل ملائمة لتمثيل تلك العلاقات بحسب ما تتطلّبه الدراسة أو مشكلة البحث.

ويعتمد التحليل الاستكشافي للبيانات على استخدام التمثيل البياني الإحصائي للبيانات (مثل المدرج التكراري، وشكل الصندوق، والأعمدة البيانية، ...، وغيرها)، إضافة لاستخدام مقاييس النزعة المركزية (ومن أهمها الوسط، والوسيط، والربيعات، ...، وغيرها)، ومقاييس التشتت (ومن أهمها الانحراف المعياري، ومعامل الاختلاف، ...، وغيرها)، إضافة لاستخدام مقاييس الارتباط بين المتغيرات.

وتجدر الإشارة هنا إلى ضرورة استخدام التحليل الاستكشافي للبيانات عند التعامل مع البيانات في أية دراسة أو مشكلة بحث حتى وإن لم يكن ذلك التحليل مطلوبا بحد ذاته، وذلك لاستكشاف سلوك هذه البيانات من جهة، واستكشاف الأخطاء أو العيوب التي قد تتعرض لها البيانات وإصلاحها إن وجدت من جهة أخرى.

وفيما يلي، سنقوم بتناول كيفية تطبيق التحليل الاستكشافي للمتغيرات الوصفية أولا يليها المتغيرات الكمية؛

3.3 التحليل الاستكشافي للمتغيرات الوصفية (EDA for Categorical Variables)

كما تمت الإشارة في البند (1.3) الخاص بتعريف أنواع البيانات، فقد تم توضيح أن المتغيرات الوصفية تكون عادة اسمية أو رتبية، ويتم عادة استخدام أشكال محددة من التمثيل البياني، أهمها الأعمدة البيانية والقطاعات البيانية، إضافة لمقاييس محدودة، مثل النسبة والمنوال، وذلك لاستكشاف سلوك هذه الأنواع من المتغيرات. إلا أنه تجدر الإشارة هنا إلى أن بعض الإحصائيين يميلون للتعامل مع المتغيرات الوصفية الرتبية التي تأخذ خمسة مستويات أو أكثر على أنها متغيرات شبه كمية، أي أنه يمكن استخدام المقاييس الكمية لدراسة سلوكها.

عند بدء التعامل مع المتغيرات الوصفية، نقوم عادة بتلخيصها فيما يعرف **بجدول التوزيع التكراري** (Frequency Distribution Table)، وهو يشبه إلى حد ما جدول التوزيع التكراري المعروف للمتغيرات الكمية حيث يتم تكوين فترات لقيم المتغير الكمي وحساب التكرارات لها، أما في حال المتغير الوصفي فإنه يتم حساب التكرارات لكل

[1] معظم الأساليب الإحصائية التقليدية التي تعتمد على تقدير المعالم أو ما يعرف **بالتقدير المعلمي** (Parametric Estimation) تضع قيودا أو فرضيات صارمة تشترط توفرها في البيانات المستخدمة في تنفيذ هذه الأساليب مثل كونها بيانات كمية لها وحدة قياس معرّفة، أو أنها تتبع توزيعا احتمالية محددا مثل التوزيع الطبيعي، أو أن تكون متغيرات الدراسة مستقلة، أو غير ذلك من الفرضيات. أما الأساليب التي تعتمد على **التقدير غير المعلمي** (Non-Parametric Estimation) فعادة ما تكون قيودها أقل.

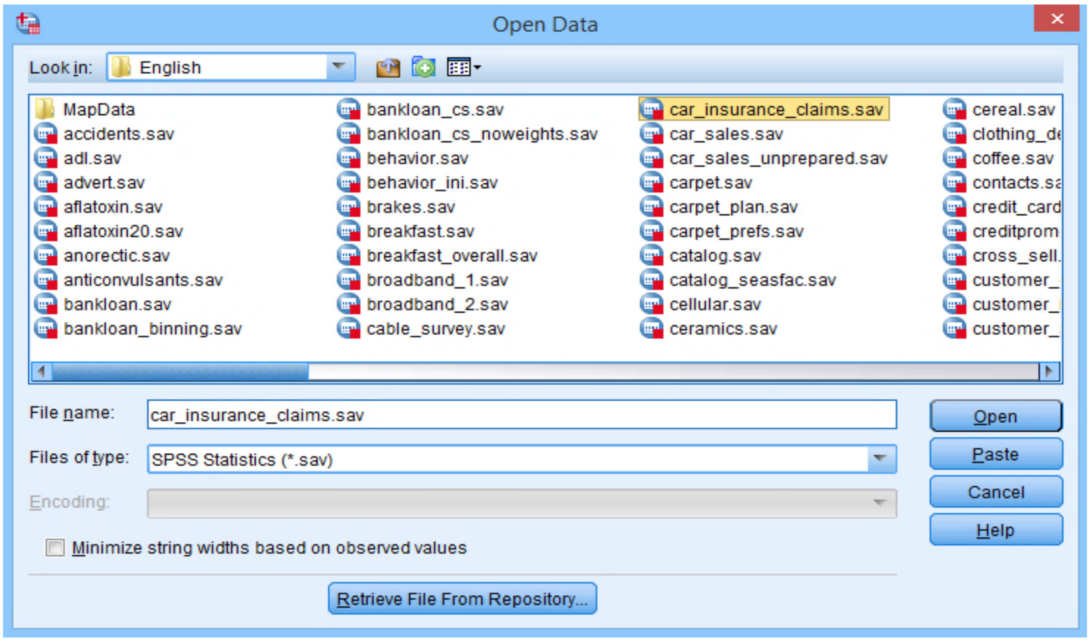
تصنيف من التصنيفات التي يأخذها هذا المتغير. ويجب الإشارة هنا إلى أن برنامج SPSS يقوم بتكوين جدول توزيع تكراري للمتغيرات عن طريق حساب التكرار لكل "قيمة" من قيم المتغير الذي يتم تعريفه بغض النظر ما إذا كانت هذه القيم كمية أو وصفية. وبالتالي، قد لا يكون من الملائم تكوين هذه الجداول للمتغيرات الكمية التي تحتوي على قيم كثيرة الاختلاف فيما بينها؛ (فمثلا، المتغير الكمي الذي يمثل درجات طلبة من 100 درجة قد يكون له جدول توزيع تكراري يضم 101 تكرار مناظر لقيم المتغير على اعتبار أن الدرجات قد تتراوح من 0 إلى 100 درجة للطالب، ومثل هذه الجداول قد لا يكون تلخيصا جيدا لهذا المتغير).

1.3.3 التوزيع التكراري للمتغيرات الوصفية (Frequency Distribution for Categorical Variables)

سنقوم الآن باستخدام ملف بيانات متوفر ضمن قاعدة البيانات المدرجة في برنامج SPSS، لتوضيح كيفية تكوين جدول التوزيع التكراري. ولفتح هذه الملفات، نقوم بالضغط على أمر ملف (File) في سطر الأوامر العلوي ثم نختار فتح (Open) ثم بيانات (Data)، فتظهر النافذة الخاصة بفتح البيانات. نقوم باختيار المجلد[1] التالي عن طريق اختيار المجلدات التسلسلية المُتضمنَة الواحد تلو الآخر؛

Program Files>IBM>SPSS>Statistics>23>Samples>English

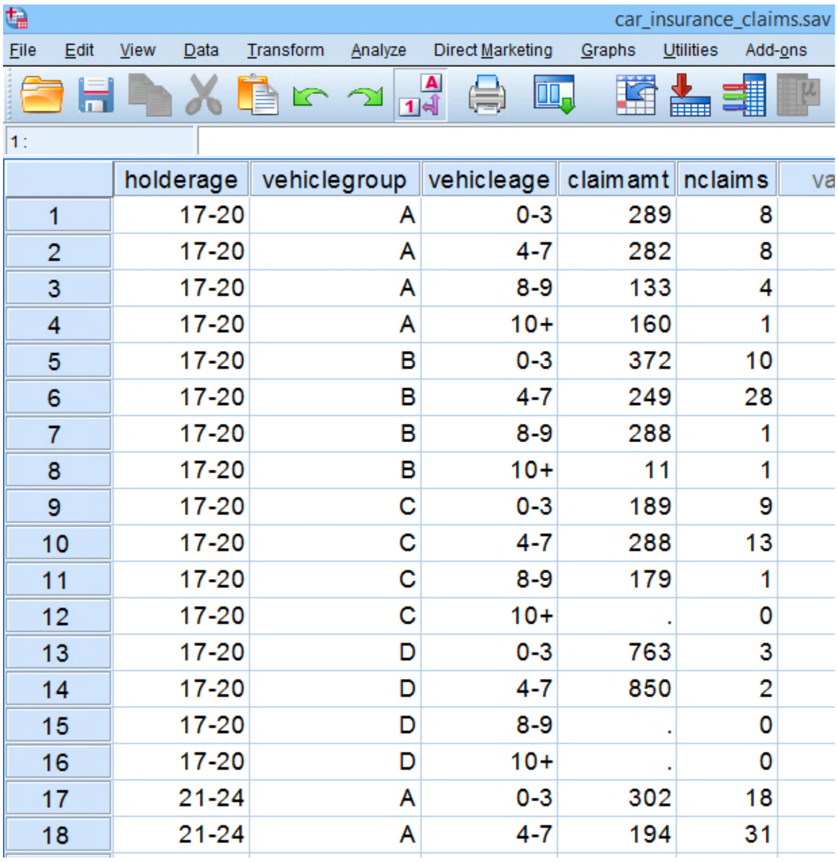
فنحصل على ملفات قاعدة البيانات في برنامج SPSS كما يظهر في الشكل (1.3). وسنقوم باختيار ملف البيانات المسمى "car_insurance_claims.sav" كما هو موضح في نفس النافذة.



شكل 1.3: نافذة ملفات قاعدة البيانات في SPSS.

[1] مجلد Program Files يكون في معظم أنظمة ويندوز ضمن المسار الأساسي وهو المسار (C:).

بعد فتح الملف، ستلاحظ أن البيانات تحتوي على خمسة متغيرات، (شكل (2.3))، تخص بيانات حول 128 شخصا في إحدى شركات التأمين على السيارات.



	holderage	vehiclegroup	vehicleage	claimamt	nclaims	va
1	17-20	A	0-3	289	8	
2	17-20	A	4-7	282	8	
3	17-20	A	8-9	133	4	
4	17-20	A	10+	160	1	
5	17-20	B	0-3	372	10	
6	17-20	B	4-7	249	28	
7	17-20	B	8-9	288	1	
8	17-20	B	10+	11	1	
9	17-20	C	0-3	189	9	
10	17-20	C	4-7	288	13	
11	17-20	C	8-9	179	1	
12	17-20	C	10+	.	0	
13	17-20	D	0-3	763	3	
14	17-20	D	4-7	850	2	
15	17-20	D	8-9	.	0	
16	17-20	D	10+	.	0	
17	21-24	A	0-3	302	18	
18	21-24	A	4-7	194	31	

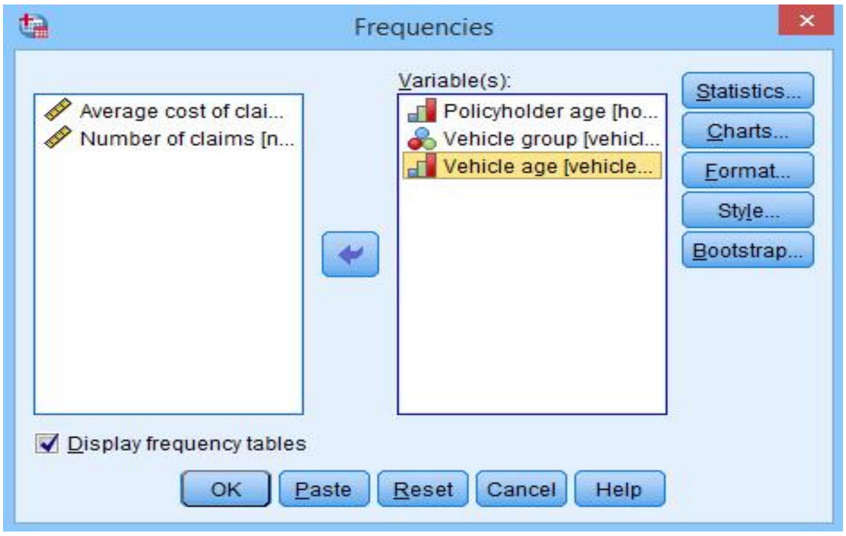
شكل 2.3: نافذة ملف البيانات "car_insurance_claims.sav" في برنامج SPSS.

وهذه المتغيرات معرّفة كالتالي؛

- holderage: عمر المؤمّن على السيارة بالسنوات، ضمن الفئات العمرية؛ (17–20)، (21–24)، ...، (60 فأكثر).
- vehiclegroup: فئة تصنيف السيارة، من ضمن الفئات؛ A، B، C، و D.
- vehicleage: عمر السيارة بالسنوات، من ضمن الفئات العمرية؛ (0–3)، (4–7)، (8–9)، و (10 فأكثر).
- claimamt: معدل قيمة التأمين المدفوعة من قِبل الشركة بالدولار.
- nclaims: عدد مبالغ التأمين المدفوعة.

ولاحظ أن المتغيرين الأول والثالث هما متغيران وصفيان رتبيان، والمتغير الثاني هو متغير وصفي اسمي، أما المتغيران الأخيران فهما كميان. وسنقوم الآن بتكوين جداول التوزيع التكراري للمتغيرات الثلاثة الأولى (الوصفية) على النحو التالي؛

نقوم من شريط الأوامر العلوي باختيار (Analyze> Descriptive Statistics> Frequencies)، فتظهر النافذة الخاصة بتنفيذ الجداول التكرارية فنقوم باختيار المتغيرات الثلاثة الأولى كما يظهر في الشكل (3.3). وفي يمين هذه النافذة، نلاحظ وجود بعض الخيارات الفرعية التي يمكن للمستخدِم استخدامها للحصول على نتائج إضافية قد يرغب بها، وسنقوم هنا بالتطرق للخيار الثاني وهو الرسم البياني (Charts)، وأما الخيار الأول فهو خاص بطلب حساب بعض المقاييس الإحصائية (مثل الوسط الحسابي والوسيط والربيعات والانحراف المعياري وغيرها من مقاييس النزعة المركزية والتشتت)، وسيتم تناول استخدام هذه المقاييس لاحقا بصورة أكثر تعمقا[1] في هذا الفصل.



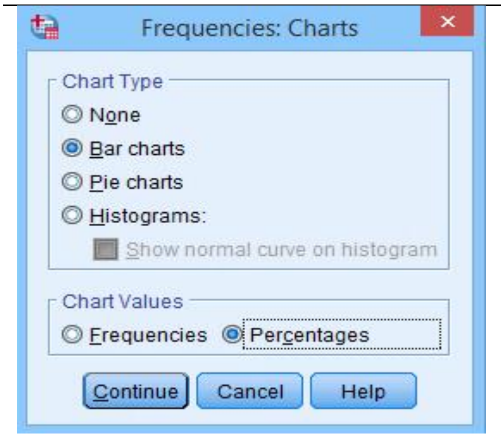
شكل 3.3: نافذة تنفيذ أمر تكرارات (Frequencies) لملف البيانات "car_insurance_claims.sav".

الآن إذا ما ضغطنا على أمر موافق (OK) في هذه النافذة فسنحصل على جداول التوزيع التكراري لكل متغير، إلا أننا في هذا المثال سنقوم باختيار تنفيذ الرسم البياني للمتغيرات للحصول على توضيح "مرئي" لتوزيع المتغيرات، أي الحصول على تمثيل بياني لما تمثله جداول التوزيع التكراري.

بالضغط على خيار الرسم البياني (Charts) في النافذة الحالية (شكل (3.3)) ستظهر نافذة فرعية كما يوضح الشكل (4.3) تشتمل على ثلاثة أنواع من الرسم هي الأعمدة البيانية (Bar Charts)، القطاعات البيانية (Pie Charts)، والمدرجات التكرارية (Histograms). وسنقوم باختيار رسم القطاعات البيانية لتمثيل المتغيرات في مثالنا.

[1] بالنسبة للخيارات الثلاثة المتبقية في يمين الشكل (3.3)، فالخياران الثالث والرابع هما للحصول على النتائج المطلوبة بترتيبات معينة وأما الخيار الخامس فهو خاص بتطبيق أسلوب إعادة المعاينة أو **البوتستراب** (Bootstrap).

وفي القســـم الســـفلي من النافذة في الشـــكل (4.3) يتوفر خيـــاران لعرض القيم على الرســـم همـــا التكرارات (Frequencies) والنسب (Percentages)، وسنختار في مثالنا هنا خيار النســـب.[1] بعد الضـــغط على اســـتمرار (Continue) في الشـــكل (4.3) ســـتختفي النافذة الفرعية وتبقى النافذة في الشـكل (3.3) فنقوم بالضـغط على موافق (OK).



شكل 4.3: نافذة اختيار نوع الرسم البياني (Charts) من نافذة تنفيذ الجدول التكراري.

ستظهر النتائج المطلوبة في نافذة النتائج (Output)، وسيتم عرضها بالترتيب التالي بحسب النسق الذي تظهر عليه في برنامج SPSS؛

أولاً: عدد المشاهدات لكل متغير (جدول (1.3)) وعدد القيم المفقودة. ونلاحظ أن عدد المشاهدات (N) هو 128 مشاهدة (المشار إليه بالفعلي (Valid)) لكل متغير من المتغيرات الثلاثة، وأنه لا توجد أي قيمة مفقودة في أي متغير (والمشار إليه بالمفقود (Missing)).

جدول 1.3: عدد المشاهدات لكل متغير في ملف البيانات "car_insurance_claims.sav".

Statistics

		Policyholder age	Vehicle group	Vehicle age
N	Valid	128	128	128
	Missing	0	0	0

ثانيا: جدول التوزيع التكراري لكل متغير، وسنقوم هنا بعرض جدول التوزيع التكراري للمتغير الأول فقط، (وهو المتغير Policyholder age)، في جدول (2.3) ويمكنك الرجوع لبقية الجداول في نافذة النتائج في برنامج SPSS.

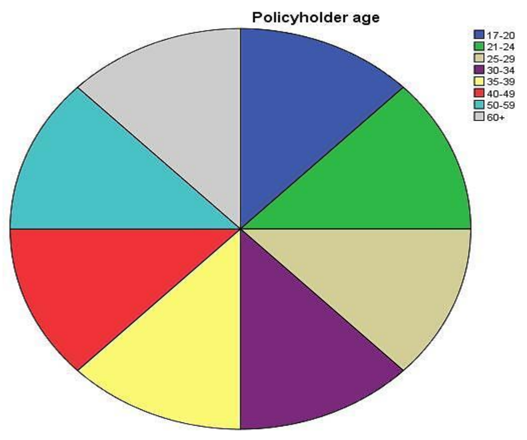
[1] في الرسم البياني في SPSS لن تظهر النسب أو القيم على القطاعات أو الأعمدة البيانية تلقائيا، بل يتم إظهارها على الشكل باستخدام نافذة محرر الرسم (Chart Editor) كما سنوضح لاحقا.

جدول 2.3: جدول التوزيع التكراري للمتغير Policyholder age في ملف البيانات "car_insurance_claims.sav".

Policyholder age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	17-20	16	12.5	12.5	12.5
	21-24	16	12.5	12.5	25.0
	25-29	16	12.5	12.5	37.5
	30-34	16	12.5	12.5	50.0
	35-39	16	12.5	12.5	62.5
	40-49	16	12.5	12.5	75.0
	50-59	16	12.5	12.5	87.5
	60+	16	12.5	12.5	100.0
	Total	128	100.0	100.0	

ويُلاحظ من الجدول (2.3) أن العامود الأول من اليسار يمثل التصنيف أو التقسيم للمتغير الوصفي (عمر المؤمّن "Policyholder age") وهو عبارة عن ثمانية تقسيمات تمثل ثمانية فئات عمرية، والعامود الثاني يمثل تكرارات كل تقسيم أي كل فئة عمرية، والعامودين الثالث والرابع يمثلان نسب الأشخاص الكلية[1] والفعلية على التوالي ضمن كل فئة عمرية. أما العامود الأخير فيمثل النسب التراكمية، أي حاصل جمع نسبة كل فئة عمرية مع ما يسبقها. ومن هذا الجدول نلاحظ أن نسب الأشخاص في كل فئة عمرية هو متساوي ويساوي 12.5%. وكذلك الحال في المتغيرين الأخرين، فإننا نلاحظ تساوي نسب السيارات في التقسيمات الأربع أي أن نسبة السيارات في كل تقسيم هو 25%، وكذلك تساوي نسب الأشخاص من حيث أعمار سياراتهم.



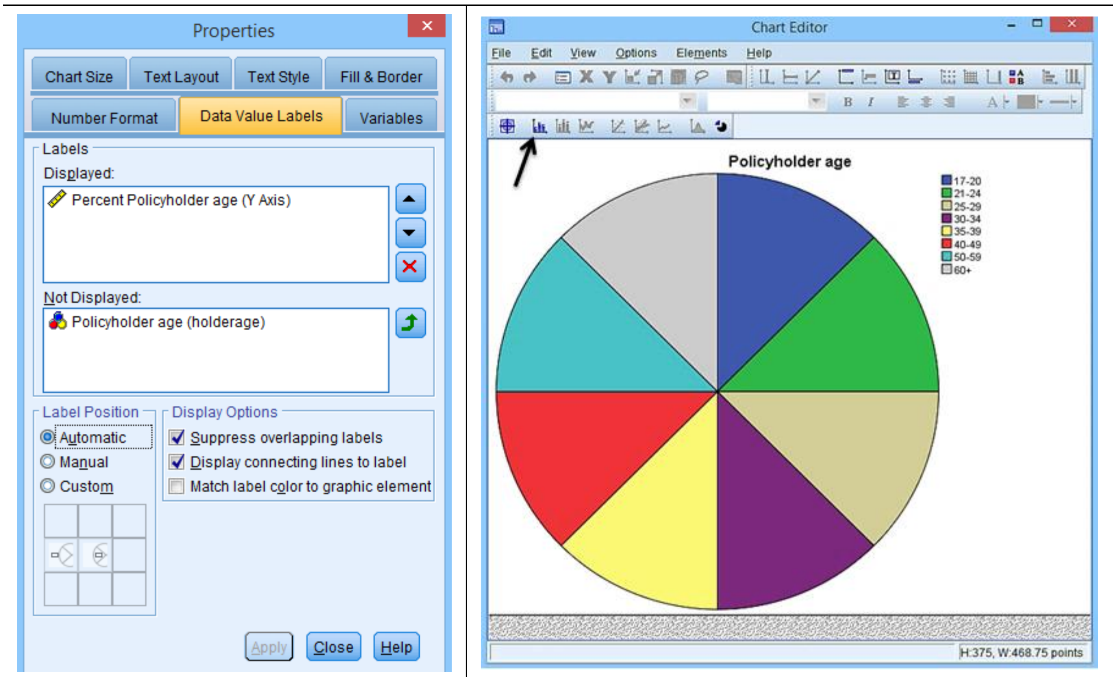
شكل 5.3: القطاعات البيانية للمتغير Policyholder age في ملف البيانات "car_insurance_claims.sav".

ثالثاً: الرسم البياني، وقد تم اختيار القطاعات البيانية في مثالنا، وسنقوم بعرض الرسم البياني للمتغير الأول فقط في الشكل (5.3) ويمكن الرجوع لبقية الرسومات في نافذة النتائج في برنامج SPSS. ويُلاحظ من الرسم البياني تساوي جميع القطاعات مما يدل على توزع عدد أو نسب أعمار الأشخاص في التقسيمات الثمانية. وكذلك الأمر بالنسبة للمتغيرين الآخرين حيث يُلاحظ تساوي توزيع النسب في القطاعات أو التقسيمات. وكما

[1] يُقصد بالنسب الكلية التي تحتوي على القيم الفعلية والمفقودة إن وجدت، وحيث أن المتغير لا يحتوي على قيم مفقودة فالنسب في العامودين الثالث والرابع تكون متساوية.

ذكرنا سابقا، فإن نسب القطاعات لم تظهر في الشكل، لذلك إذا ما أردنا إظهارها على القطاعات أو على أي رسم بياني نقوم بالخطوات التالية:

في نافذة النتائج (Output) في SPSS للمثال السابق، قم بالنقر المزدوج (بالزر الإيسر للفأرة) على رسم القطاعات البيانية للمتغير (Policyholder age) فتظهر نافذة فرعية هي نافذة محرر الرسم (Chart Editor) كما يوضح الشكل (6.3 أ).



(ب) (أ)

شكل 6.3: (أ) نافذة محرر الرسم (Chart Editor) لرسم القطاعات البيانية للمتغير (Policyholder age) و(ب) نافذة خواص محرر الرسم الخاصة بها.

قم بالضغط على الأيقونة المشار لها بالسهم في الشكل، والتي تمثل إظهار وسم البيانات (Show Data Label) والتي تمثل في مثالنا نسب القطاعات، عندها ستظهر نافذة فرعية أخرى، كما يُشاهد في الشكل المجاور، الشكل (6.3 ب)، وفي نفس الوقت ستظهر تلك النسب على القطاعات في نافذة النتائج في SPSS. قم بعدها بالضغط على أمر الإغلاق (Close) في النافذة الأخيرة وكذلك يمكنك إغلاق نافذة محرر الرسم إذا لم يكن هنالك تعديلات تريد تنفيذها. وستلاحظ بعد ذلك ظهور نسب الأعمار للمتغير المطلوب على القطاعات البيانية في نافذة النتائج.

وفي كلا النافذتين، (الشكل (6.3 أ، ب))، ستلاحظ وجود العديد من الخيارات التي يمكنك استخدامها للتعديل على الرسم مثل حجم الرسم، حجم ونوع ولون الخطوط في الرسم، إضافة عنوان للرسم، تعديل أسماء محاور الرسم، تغيير مواضع وسم البيانات، تغيير الألوان في الرسم، وغيرها من الخيارات التي سنتناول البعض منها لاحقا.

ملاحظة:

في نافذة محرر الرسم، عند النقر المزدوج على أي جزء من الرسم ستظهر نافذة خواص ذلك الجزء بالتحديد ويمكنك من خلالها إجراء التعديلات المطلوبة.

2.3.3 الأعمدة البيانية لجداول التوزيع التكراري (Bar plots of Frequency Distribution Tables)

تشبه رسومات الأعمدة البيانية للمتغيرات الوصفية، من حيث الشكل العام، رسومات المدرج التكراري والتي تستخدم لتمثيل التوزيع التكراري للمتغيرات الكمية. إلا أن الأعمدة البيانية غالبا ما تستخدم لتمثيل وعرض البيانات الوصفية غير الرقمية باستخدام تكراراتها. ويتم على الرسم توزيع مستويات أو تصنيفات المتغير الوصفي المختلفة على أحد المحاور، (المحور الأفقي عادة)، وتكرار الحدوث أو النسبة على المحور الآخر. وفي هذه الفقرة، سيتم من خلال المثال القادم شرح طريقة إدخال البيانات للمتغير الوصفي التي تكون مجهزة مسبقا بصورة جدول توزيع تكراري كما هو سائد بشكل كبير في أوساط المتعاملين بالتحليل الإحصائي للبيانات.

البيانات في الجدول التالي، (الجدول (3.3))، تمثل أعداد الطلبة الذكور والإناث في خمسة أقسام في كلية العلوم بجامعة بنغازي لعام 2012. وسيتم استخدام الأعمدة البيانية "لتحويل" محتويات الجدول إلى "صورة" معبرة عما تمثله هذه البيانات.

جدول 3.3: توزيع أعداد الطلبة الذكور والإناث في أقسام كلية العلوم بجامعة بنغازي.

	القسم				
النوع ↓	**رياضيات**	**إحصاء**	**فيزياء**	**كيمياء**	**نبات**
ذكور	190	120	170	510	470
إناث	280	310	240	720	960
المجموع	470	430	410	1230	1430

ستكون أول خطوة هنا هي إدخال البيانات في ملف بيانات جديد في SPSS وليكن باسم "أقسام كلية العلوم"، كما يظهر في الشكل (7.3). ولاحظ هنا أنه تم تعريف ثلاثة متغيرات هي "الطلبة" وهو متغير كمي يمثل أعداد الطلبة، و"النوع" وهو متغير وصفي اسمي يمثل نوع الطالب ذكر أو أنثى، و"القسم" وهو أيضا متغير وصفي اسمي يمثل القسم الذي ينتمي إليه الطالب من بين خمسة أقسام.

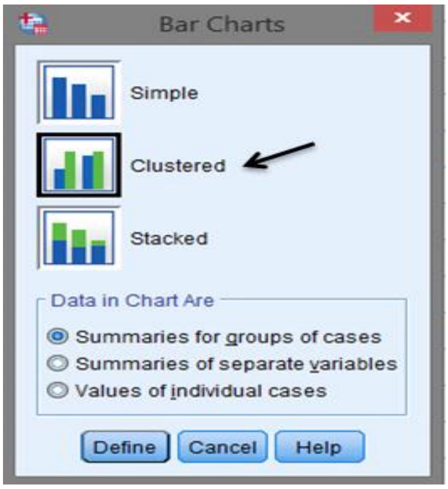
وننوه هنا بأنه قد تكوين ملف البيانات "أقسام كلية العلوم" من الجدول السابق، (جدول (3.3))، عن طريق إدخال قيم الصف الأول (190، 120، ...، 470) متبوعة بقيم الصف الثاني (280، 310، ...، 960) كقيم للمتغير الأول وهو "الطلبة"، ثم تم إدخال تصنيف نوع الطالب والقسم الذي يتبع له، (وهما المتغيران "النوع" و"القسم")، بحسب ترتيب إدخال أعداد الطلبة، فمثلا القيمة 190 تمثل الطلبة الذكور في قسم الرياضيات، والقيمة 120 تمثل الطلبة الذكور في قسم الإحصاء، وهكذا.



	الطلبة	النوع	القسم
1	190	ذكور	رياضيات
2	120	ذكور	إحصاء
3	170	ذكور	فيزياء
4	510	ذكور	كيمياء
5	470	ذكور	نبات
6	280	إناث	رياضيات
7	310	إناث	إحصاء
8	240	إناث	فيزياء
9	720	إناث	كيمياء
10	960	إناث	نبات

شكل 7.3: عرض بيانات ملف "أقسام كلية العلوم".

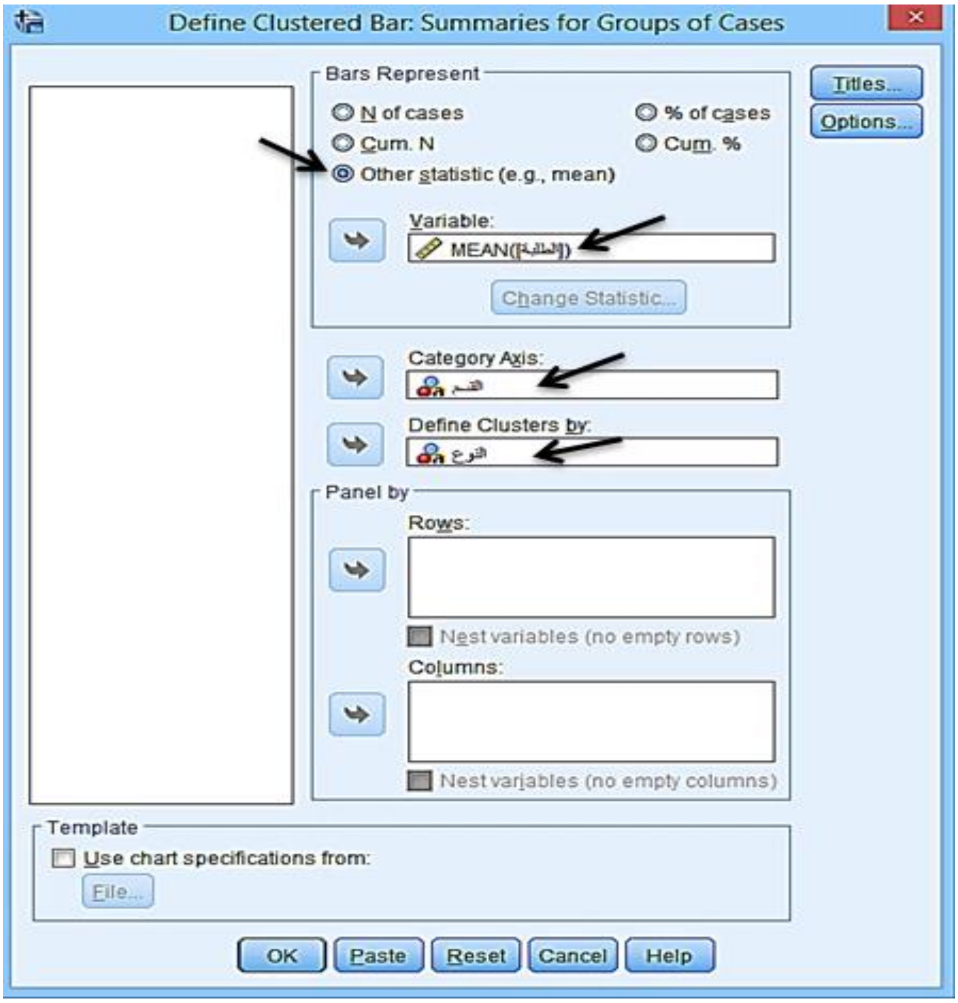
الآن سنقوم بتمثيل البيانات في ملف البيانات "أقسام كلية العلوم" باستخدام الأعمدة البيانية المصنفة (Clustered Bar charts)، والذي سيتم من خلاله عرض أعداد الطلبة بحسب نوع الطالب والقسم التابع له. في شريط الأدوات العلوي في برنامج SPSS قم باختيار Graphs > Legacy Dialogs > Bar فتظهر النافذة الفرعية التالية التي تتيح اختيار نوع الأعمدة البيانية؛ (شكل (8.3)).



شكل 8.3: نافذة اختيار نوع الأعمدة البيانية (Clustered).

قم باختيار الأعمدة البيانية المصنفة كما هو مشار إليه بالسهم (Clustered) ثم اضغط تعريف (Define). ستظهر بعد ذلك نافذة أخرى، (شكل (9.3))، والتي سيتم من خلالها تعريف ترتيب متغيرات التصنيف، (والتي هي في مثالنا الحالي تمثل نوع الطالب والقسم التابع له).

قم في تلك النافذة بالنقر على الخيار (Other statistic) والمشار إليه بالسهم، ثم عرف المتغيرات الثلاثة في ملف البيانات "أقسام كلية العلوم" بنفس الترتيب الموضح في الشكل (9.3) والمشار إليه بالأسهم، حيث أن المتغير الكمي "الطلبة" سيكون هو المتغير الأساسي والمتغير "القسم" سيمثل التصنيف على المحور الأفقي، والمتغير "النوع" سيكون ضمن تصنيف الأقسام.

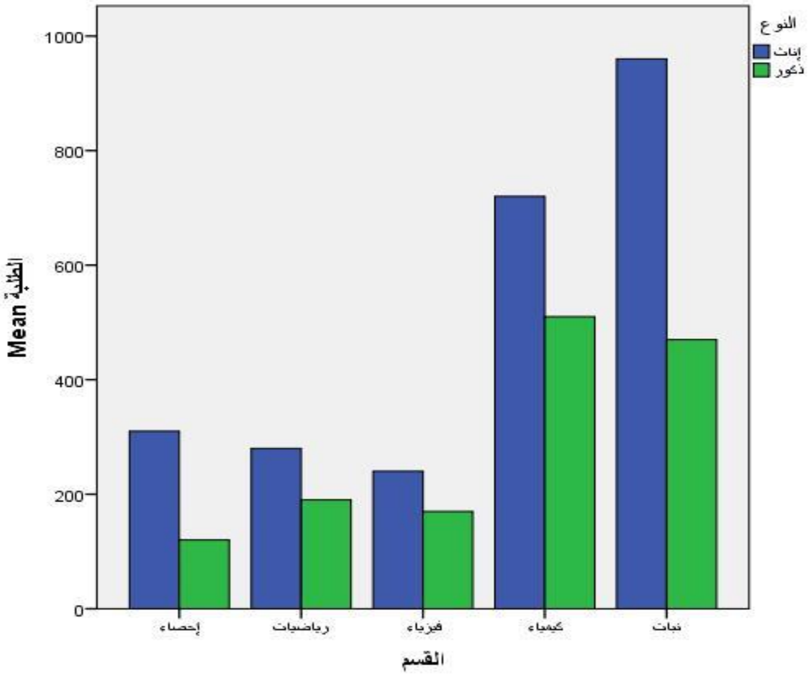


شكل 9.3: نافذة تعريف المتغيرات للأعمدة البيانية المصنفة لبيانات ملف "أقسام كلية العلوم".

بعد الانتهاء من تعريف المتغيرات[1] قم بالضغط على موافق (OK) في أسفل النافذة للتنفيذ. سنحصل مباشرة بعد ذلك على الأعمدة البيانية المصنفة لأعداد الطلبة في نافذة المخرجات، كما هو في الشكل (10.3). ويمكن الوقوف على عدة ملاحظات من شكل الأعمدة البيانية من أهمها؛

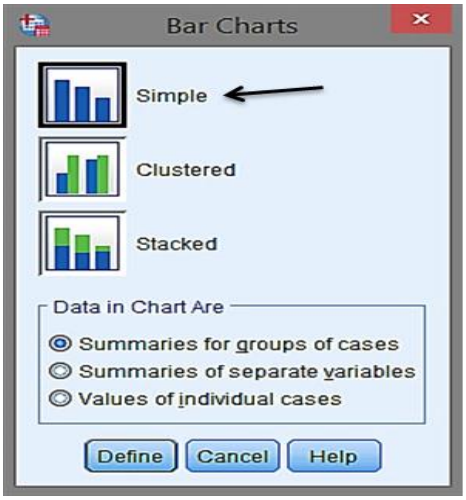
- تميز قسمي النبات والكيمياء بوجود أعداد أكبر من الطلبة من باقي الأقسام.
- تفوق أعداد الطالبات على أعداد الطلبة في جميع أقسام الكلية وخاصة في قسمي النبات والكيمياء.
- أقل عدد للطلبة هو في قسم الفيزياء.

[1] يمكنك إدراج عناوين أو ملاحظات خاصة بك ضمن الرسم عن طريق الدخول إلى العناوين (Titles) في أعلى النافذة إلى اليمين. كما يمكنك أيضا الحصول على بعض الخيارات الإضافية على الرسم، (مثل فترات الثقة وغيرها) بالدخول إلى الخيارات (Options).



شكل 10.3: الأعمدة البيانية المصنفة للطلبة بحسب النوع والقسم لبيانات ملف البيانات "أقسام كلية العلوم"

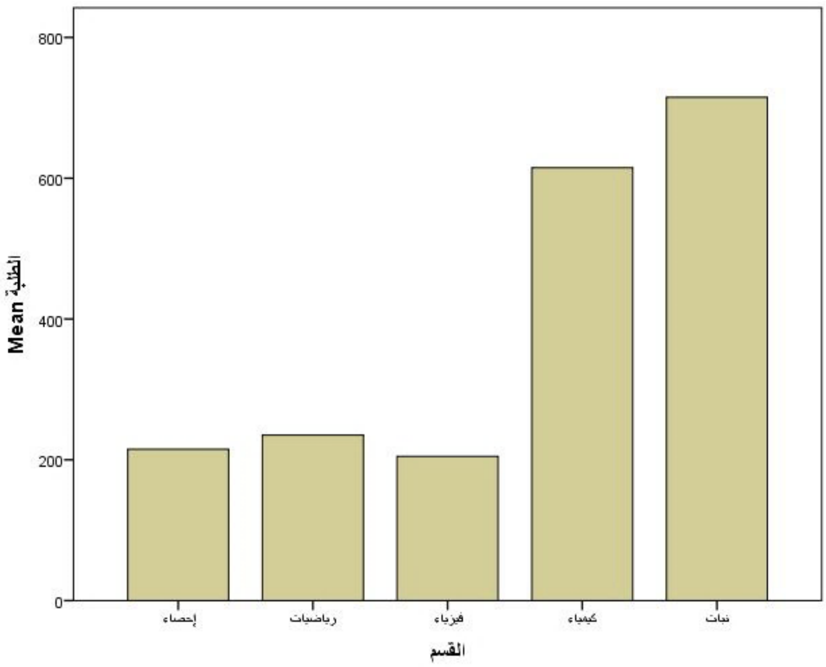
من جديد، يمكننا ولنفس المثال، (ملف البيانات "أقسام كلية العلوم")، استخدام الأعمدة البيانية لتمثيل أعداد الطلبة في الأقسام بدون تصنيفهم إلى ذكور وإناث. ولتنفيذ ذلك، قم في شريط الأدوات العلوي قم باختيار Graphs>Legacy Dialogs>Bar فتظهر النافذة الفرعية التي نراها في الشكل (11.3). قم باختيار الأعمدة البيانية البسيطة (Simple) كما هو مشار إليه بالسهم ثم اضغط تعريف (Define).



شكل 11.3: نافذة اختيار نوع الأعمدة البيانية البسيطة (Simple).

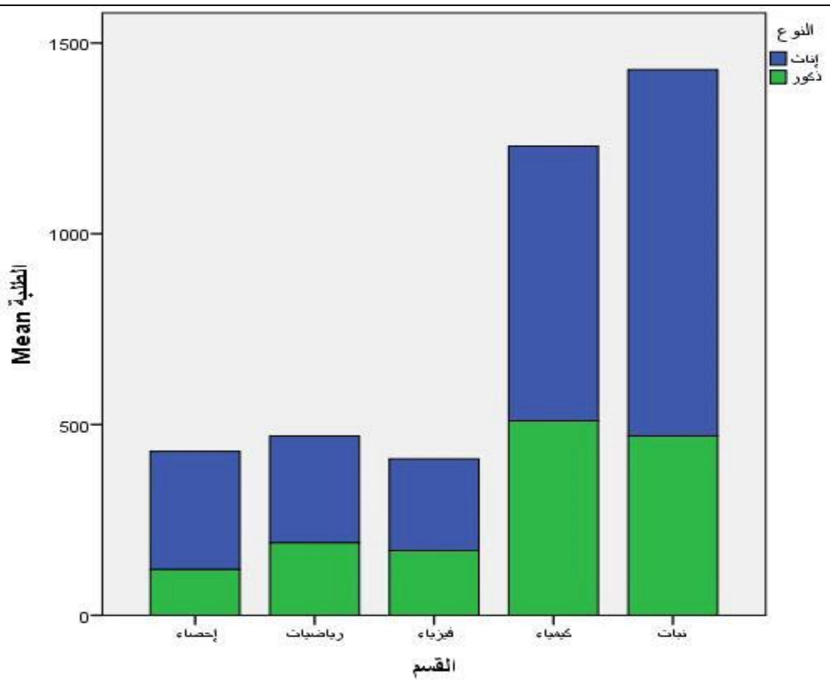
ستظهر بعد ذلك نافذة تعريف ترتيب المتغيرات التصنيف، كما هو في الشكل (9.3) السابق، قم في تلك النافذة بالنقر على الخيار إحصاءات أخرى (Other statistic) والمشار إليه بالسهم، ثم عرف المتغيران "الطلبة" و"القسم" بنفس الترتيب الموضح والمشار إليه بالأسهم ولكن بدون تعريف المتغير الثالث وهو "النوع"، حيث أن المتغير الكمي "الطلبة" سيكون هو المتغير الأساسي والمتغير "القسم" سيمثل التصنيف على المحور الأفقي.

بعد الانتهاء من تعريف المتغيرات قم بالضغط على موافق (OK) في أسفل النافذة وستحصل بعد ذلك على الأعمدة البيانية لأعداد الطلبة بحسب القسم فقط في نافذة المخرجات، كما هو في الشكل (12.3).



شكل 12.3: الأعمدة البيانية المصنفة للطلبة بحسب القسم لبيانات ملف البيانات "أقسام كلية العلوم".

ويمكننا من شكل (12.3) ملاحظة وجود أعداد أكبر من الطلبة في قسمي النبات والكيمياء، وكذلك انخفاض عدد طلبة قسم الفيزياء مقارنة بباقي أقسام الكلية، وهذه الملاحظات أمكن رؤيتها من الشكل (10.3) السابق أيضا. ونشير هنا إلى أنه إذا ما تم اختيار الخيار الثالث (Stacked) في الشكل (8.3) فإننا سنحصل على أعمدة بيانية مصنفة بحسب القسم والنوع كما يظهر في الشكل (13.3).



شكل 13.3: الأعمدة البيانية المصنفة للطلبة بحسب القسم والنوع لبيانات ملف "أقسام كلية العلوم" باستخدام خيار (Stacked).

<u>ملاحظة</u>:

يمكن استخدام القطاعات البيانية، كما أوضحنا سابقا، مع بيانات المثال السابق لتمثيل أعداد الطلبة بحسب القسم (أو النوع) وسنحصل على نفس الملاحظات ولكن بتمثيل بياني مختلف.

3.3.3 الأعمدة البيانية للمتغيرات الوهمية (Bar plots for Dummy Variables)

في بعض الأحيان قد نحتاج لمقارنة تغير أو توزع البيانات ضمن عدة متغيرات وصفية تأخذ طابعا خاصا مثل أن يتم المقارنة بين آراء الناس حول موضوع معين بحيث يمكن للشخص المستجيب أن يعطي أكثر من إجابة واحدة بنعم أو لا لنفس السؤال، أي أن يكون للشخص أكثر من خيار لكل إجابة بنعم أو لا، في هذه الحالة يتم تعريف مجموعة من المتغيرات، والتي تسمى عندها بالمتغيرات الوهمية، بحيث أن كل متغير من هذه المتغيرات يمثل استجابة الشخص بنعم أو لا لكل خيار. وقد سبق التعرض للمتغير الوهمي في الفصل الثاني، (البند (4.2))، وسوف نقدم مثال آخر هنا للمزيد من التوضيح؛

قامت إحدى مراكز الصحية المختصة بعلاج السمنة بتوزيع استبيان على مجموعة من الأشخاص لمعرفة آرائهم حول وجباتهم المفضلة من المطاعم المختلفة، فكانت النتائج كما هو موضح في الجدول التالي، (جدول (4.3)) والرقم "1" يدل على اختيار الشخص لنوع الطعام المفضل، فمثلا؛ الشخص الأول في الجدول يفضل تناول الدجاج واللحوم الحمراء والمثلجات، وهكذا.

جدول 4.3: آراء عينة من الأشخاص حول مأكولاتهم المفضلة.

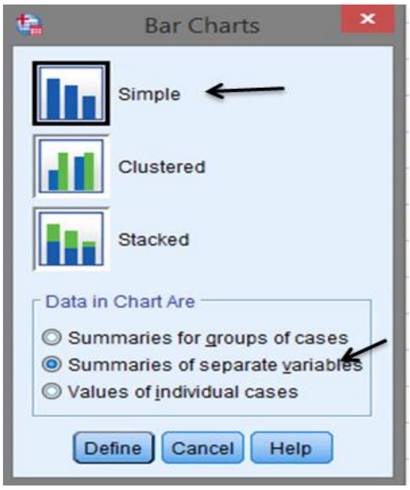
	بيتزا	دجاج	سلطات	لحوم حمراء	مثلجات
1		1		1	1
2	1			1	1
3					1
4	1	1		1	
5	1	1	1	1	1
6	1			1	1
7					1
8	1			1	1
9			1		1
10			1	1	1
11	1	1			
12				1	1
13	1		1	1	1
14		1			
15		1		1	1
16				1	1
17	1			1	1
18		1			1
19	1	1			1
20	1		1		1

وبعد إدخال هذه البيانات في ملف جديد في برنامج SPSS، وليكن باسم "المأكولات المفضلة"، كما نشاهد في الشكل (14.3) نقوم بالخطوات التالية؛

	بيتزا	دجاج	سلطات	لحوم_حمراء	مثلجات
1	.	1	.	1	1
2	1	.	.	1	1
3	.	.	.	.	1
4	1	1	.	1	.
5	1	1	1	1	1
6	1	.	.	1	1
7	.	.	.	.	1
8	1	.	.	1	1
9	.	.	1	.	1
10	.	.	1	1	1
11	1	1	.	.	.
12	.	.	.	1	1
13	1	.	1	1	1
14	.	1	.	.	.
15	.	1	.	1	1
16	.	.	.	1	1
17	1			1	1

شكل 14.3: عرض بيانات ملف "المأكولات المفضلة".

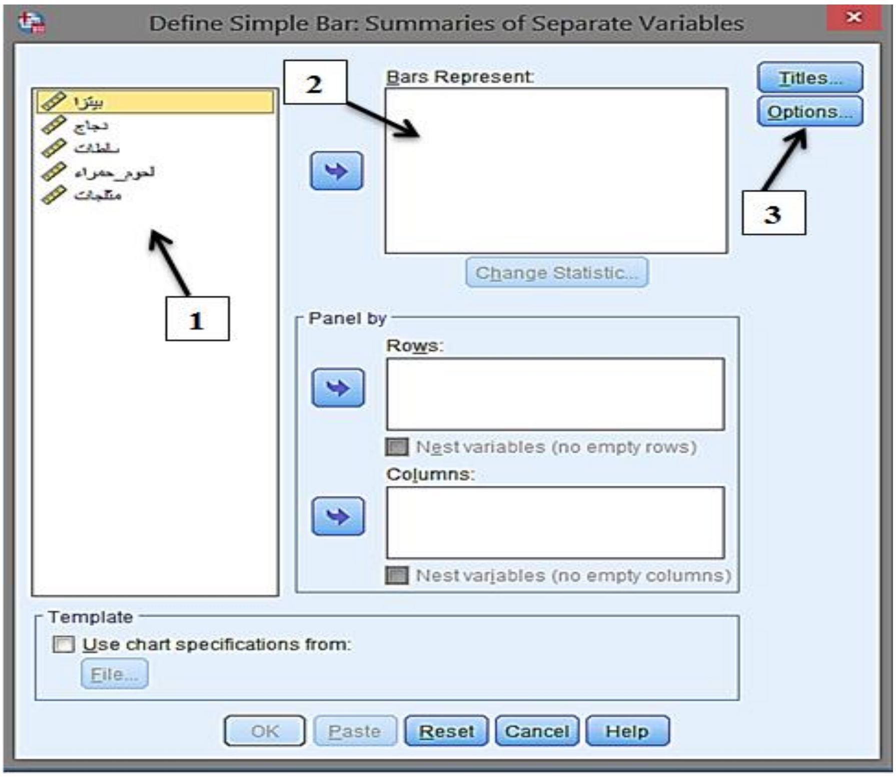
في شريط الأدوات العلوي قم باختيار Graphs>Legacy Dialogs>Bar وبعد ظهور النافذة الفرعية الخاصة باختيار نوع الأعمدة البيانية، (شكل (15.3))، قم بإبقاء اختيار الأعمدة البسيطة (Simple) ثم قم بالنقر على الخيار الثاني في الجزء الأسفل من النافذة والمشار إليه بالسهم السفلي لاختيار التعامل مع عدة متغيرات.



شكل 15.3: نافذة اختيار تمثيل الأعمدة البيانية لأكثر من متغير.

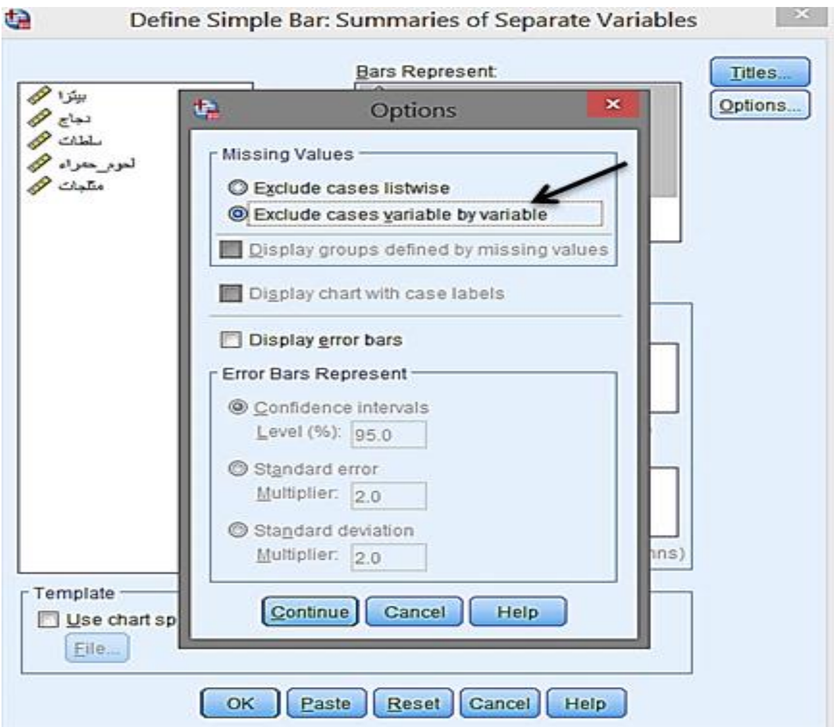
اضغط تعريف (Define) فتظهر نافذة جديدة لتعريف المتغيرات المطلوب تمثيلها باستخدام الأعمدة البيانية كما يوضح الشكل التالي، (شكل (16.3)).

في هذه النافذة، قم باختيار كل المتغيرات الخمسة كما هو مشار إليه في السهم رقم 1، ثم قم بنقلها إلى الخانة المشار إليها بالسهم رقم 2، (وهي خانة المتغيرات التي ستمثلها الأعمدة البيانية)، بعد ذلك انقر على زر الخيارات (Options) الذي يشير إليه السهم رقم 3 فتظهر نافذة فرعية جديدة كما يوضح الشكل (17.3).



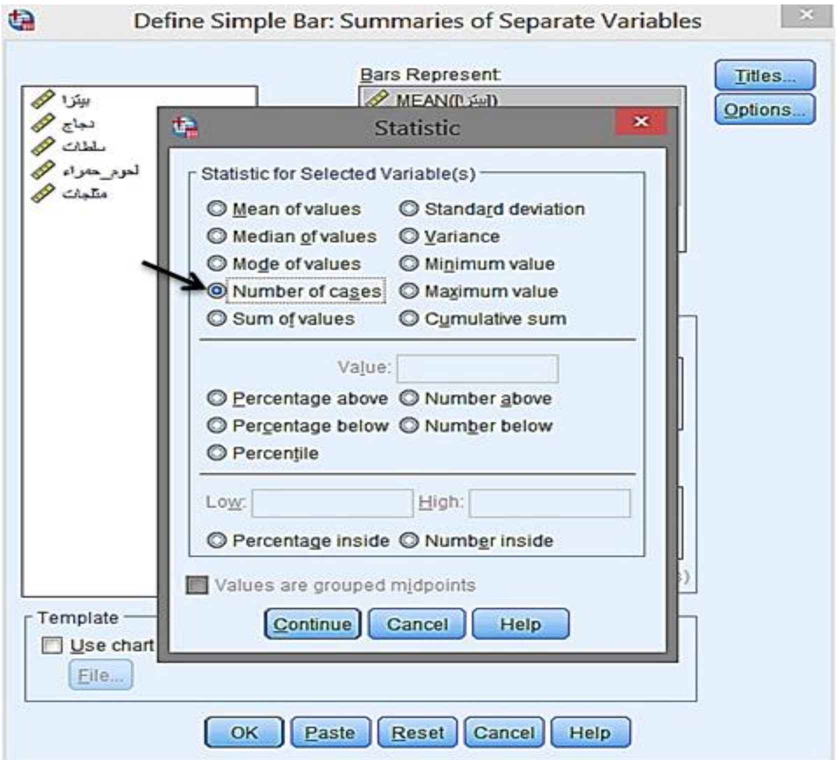
شكل 16.3: نافذة اختيار المتغيرات في تمثيل الأعمدة البيانية المتعددة.

في نافذة الشكل (17.3)، قم بالضغط على الخيار المشار إليه بالسهم، (والخاص باستثناء القيم المفقودة ضمن كل متغير وليس باستثناء الصف)، ثم اضغط استمرار (Continue).



شكل 17.3: نافذة تعديل اختيار القيم المفقودة ضمن المتغيرات.

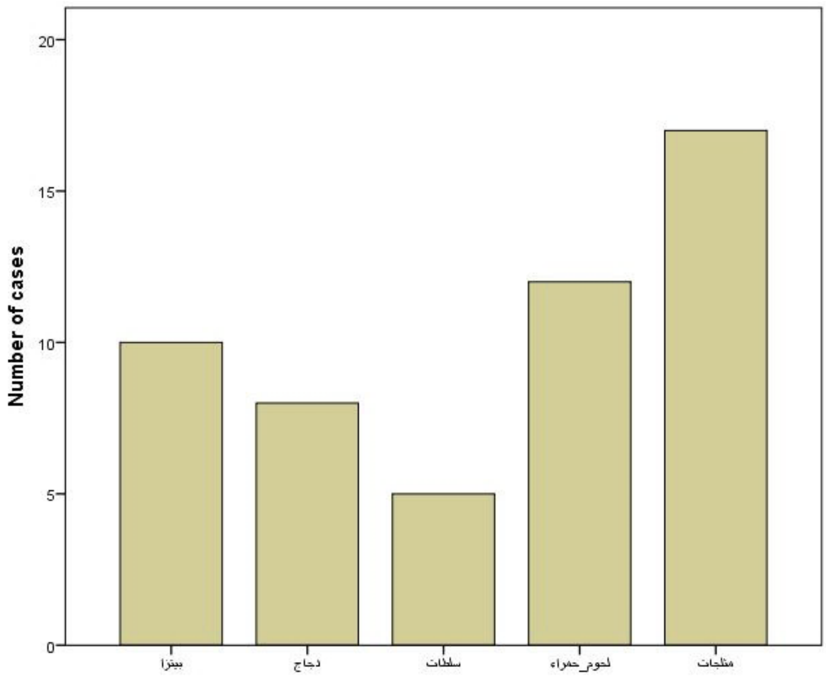
ستعود الآن للنافذة السابقة (شكل (16.3) السابق)، قم عندها بالضغط على خيار تغيير الإحصاءة (Change Statistic) والموجود أسفل الخانة التي تم نقل المتغيرات إليها، وستظهر عندها نافذة فرعية جديدة كما يظهر في الشكل (18.3).



شكل 18.3: نافذة اختيار الإحصاءة المطلوبة للمتغيرات في تمثيل الأعمدة البيانية.

قم في هذه النافذة باختيار الإحصاءة الخاصة بعدد المشاهدات لكل متغير، وهي في مثالنا الحالي ستكون عبارة عن مجموع آراء الناس أي مجموع الرقم 1 لكل متغير، بعد ذلك اضغط استمرار وستختفي النافذة الفرعية، اضغط موافق (OK) للتنفيذ. سنحصل بعدها على الأعمدة البيانية المطلوبة في نافذة المخرجات، وهذه الأعمدة موضحة في الشكل (19.3).

ومن هذه الأعمدة يمكن بوضوح القول بأن الأشخاص الذين تم استبيانهم يفضلون تناول المثلجات كرغبة أولى يليها اللحوم الحمراء والبيتزا.



شكل 19.3: الأعمدة البيانية الممثلة لآراء الناس في الطعام المفضل لبيانات ملف "المأكولات المفضلة".

4.3 التحليل الاستكشافي للمتغيرات الكمية (EDA for Quantitative Variables)

كما وضحنا في البند الخاص بأنواع البيانات، فإن البيانات الكمية تنقسم إلى فئوية ونسبية، إلا أن هذا التقسيم لن يتم أخذه بالاعتبار من الناحية العملية إلا عند التعليق على المعلومات المستخلصة من البيانات كما سنلاحظ من خلال الأمثلة التي سنتناولها في هذا البند.

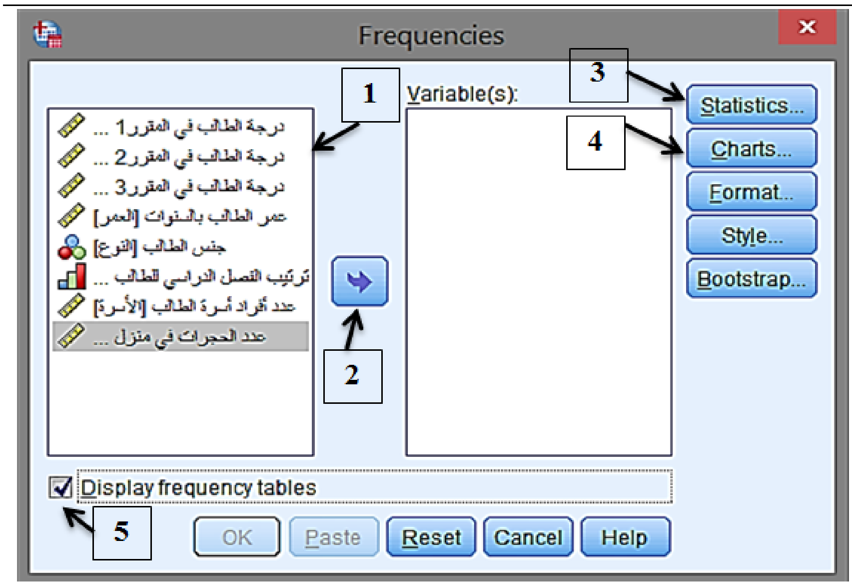
وتوجد عدة طرق ومقاييس لاستكشاف البيانات الكمية وهي تندرج في مفهوم الإحصاء الوصفي التقليدي تحت مسميات مقاييس **النزعة المركزية** (Central Tendency)، وهي تشمل أهم وأشهر المقاييس الإحصائية مثل الوسط الحسابي والوسيط والمنوال وغيرها، ومقاييس **التشتت** أو **الاختلاف** (Dispersion). وأهمها المدى والانحراف المعياري والتباين وغيرها. وكما هو الحال مع البيانات الوصفية، فإن التمثيل البياني الإحصائي يُعد أداة هامة جدا تستخدم لتوضيح واستكشاف طبيعة وتوزيع البيانات كما سنرى.

من جديد، نذكر هنا بأنه يمكن التعامل مع أية متغيرات، كمية كانت أو وصفية، بشكل مفرد (التحليل الأحادي)، أو متعدد، أي التعامل مع عدة متغيرات في آن واحد، (التحليل المتعدد). والأساليب الإحصائية التي تُستخدم مع التحليل الأحادي تعني أنه سيتم حساب تلك المقاييس لكل متغير على حدة، حتى وإن كانت قاعدة البيانات متكونة من متغيرات عديدة. وسنقوم بداية بحساب أهم المقاييس الإحصائية لاستكشاف البيانات من خلال الأمثلة التالية.

باستخدام ملف البيانات "بيانات الطلبة"، (الذي تم تكوينه في الفصل الثاني)، والذي يمثل البيانات المتعلقة بـ 35 طالبا جامعيا في إحدى الأقسام العلمية، حيث تمثل المتغيرات فيه؛

مقرر1: درجة الطالب (من 100 درجة) في المقرر1، مقرر2: درجة الطالب في المقرر2، مقرر3: درجة الطالب في المقرر3، العمر: عمر الطالب بالسنوات، النوع: جنس الطالب (1= ذكر، و2= أنثى)، الفصل: ترتيب الفصل الدراسي للطالب (المرحلة)، الأسرة: عدد أفراد أسرة الطالب، والمنزل: عدد الحجرات في منزل الطالب.

سنقوم بحساب بعض المقاييس الإحصائية الهامة والوقوف على تفسيرها. فنذهب لشريط الأدوات العلوي في برنامج SPSS ونختار التالي: Analyze>Descriptive Statistics>Frequencies فتظهر نافذة الإدخال كما هو مبين في الشكل التالي، (شكل (20.3))؛



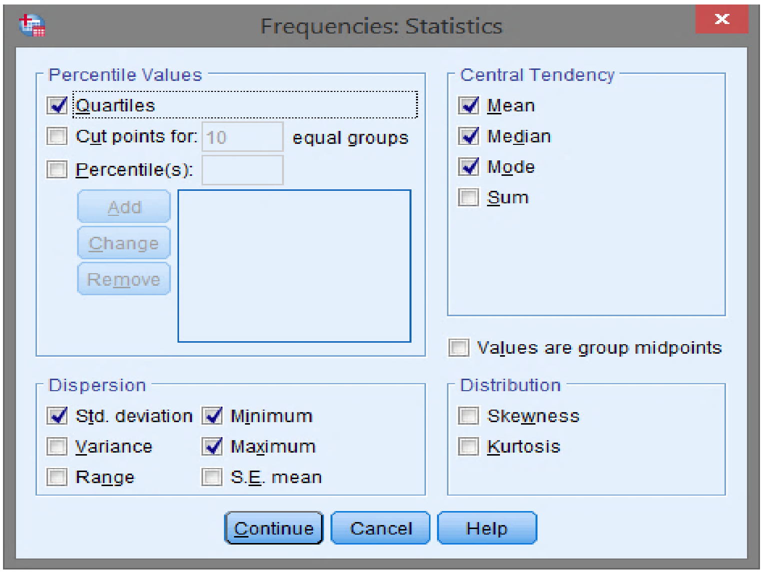
شكل 20.3: نافذة الإدخال لأمر التكرارات (Frequencies).

في هذه النافذة، يمكن للمستخدم اختيار المتغير أو المتغيرات التي يرغب في حساب المقاييس الإحصائية والرسومات البيانية لها من المربع الإيسر المشار إليه بالسهم (1)، وسنقوم في مثالنا هذا باختيار جميع[1] المتغيرات الثمانية المتوفرة. وبعد اختيار كل المتغيرات نقوم بنقلها باستخدام السهم المشار إليه بالرقم (2)، إلى مربع المتغيرات (Variables). الآن سنقوم باختيار المقاييس التي نرغب بحسابها من خيار الإحصاءات (Statistics) المشار إليه بالسهم (3).

عندها ستظهر نافذة فرعية جديدة، (الشكل (21.3))، قم في هذه النافذة الفرعية باختيار المقاييس المشار إليها بالعلامة (√) وهي **الوسط الحسابي** (Mean)، **الوسيط** (Median)، **المنوال** (Mode)، **الربيعات** (Quartiles)، **الانحراف المعياري** (Std. Deviation)، **القيمة الصغرى** (Minimum)، و**القيمة الكبرى** (Maximum). أما

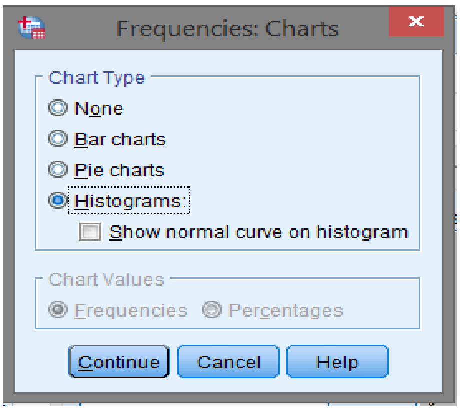
[1] لاختيار جميع المتغيرات مرة واحدة يمكن بعد النقر بالفأرة على المربع الإيسر الضغط على زر CTRL في لوحة المفاتيح والحرف A في آن واحد فيتم اختيار كافة المتغيرات.

المقاييس الأخرى المتوفرة ضمن الخيارات، (مثل **المجموع** (Sum) و**التباين** (Variance) وغيرها)، فليس هنالك حاجة لاستخدامها حاليا في هذا المثال وسنتطرق لها لاحقا بحسب الحاجة.



شكل 21.3: نافذة اختيار المقاييس الإحصائية ضمن نافذة التكرارات (Frequencies).

بعد اختيار المقاييس المطلوبة اضغط استمرار (Continue) فيتم العودة للنافذة السابقة (شكل (20.3))، ونقوم في تلك النافذة بالضغط على خيار الرسومات البيانية (Charts) المشار إليه بالسهم (4) لاختيار الرسم البياني المطلوب. ستظهر عندئذ نافذة فرعية أخرى، (شكل (22.3))، والتي يمكن من خلالها اختيار عدم استخدام أي رسم بياني (None)، أو اختيار الأعمدة البيانية (Bar Charts)، أو اختيار القطاعات البيانية (Pie Charts)، أو اختيار المدرج التكراري (Histogram)، وسنقوم هنا باختيار المدرج التكراري، والذي سيتم تفسيره في البند القادم.



شكل 22.3: نافذة اختيار الرسم البياني ضمن نافذة التكرارات (Frequencies).

ونشير هنا في نفس النافذة إلى إمكانية اختيار إظهار منحنى التوزيع الطبيعي على الرسم البياني للمدرج التكراري، إلا أننا لن نقوم باستخدامه حاليا حيث أننا سنفرد له تعليقا مطولا عند تناول موضوع التوزيع الطبيعي.

ملاحظة:

إذا ما تم اختيار الأعمدة البيانية أو القطاعات البيانية في الشكل (22.3) فإنه يمكن اختيار عرض التكرارات (Frequencies) أو النسب (Percentages) على الرسم من الخيار أسفل النافذة.

بعد اختيار الرسم المطلوب اضغط على استمرار (Continue) فيتم العودة للنافذة السابقة (شكل (20.3)). وإذا ما قمنا في تلك النافذة بإبقاء اختيار عرض الجداول التكرارية (Display Frequency Tables) المشار إليه بالسهم رقم (5)، فإن سنحصل عليها مع المقاييس والرسومات المطلوبة، إلا أننا في هذا المثال لن نكون بحاجة إليها لذلك سنقوم بإلغاء اختيار عرضها. نقوم بعد ذلك بالضغط على موافق (OK) في نافذة الشكل (20.3) فنحصل على النتائج المطلوبة تباعا في نافذة المخرجات.

النتيجة الأولى، بحسب ترتيب العرض في نافذة المخرجات ستكون المقاييس الإحصائية المطلوبة لكافة المتغيرات الثمانية كما يوضح الشكل (23.3). في العامود الأول إلى اليسار، توجد أسماء المقاييس بالترتيب التالي؛ عدد المشاهدات (N Valid) وهو 35 مشاهدة، عدد القيم المفقودة (Missing) ويساوي صفرا لعدم وجود قيم مفقودة في البيانات، المتوسط، الوسيط، المنوال، الانحراف المعياري، القيمة الصغرى، القيمة الكبرى، الربيعات: الربيع الأول، الثاني، والثالث، (ولاحظ أنها معروضة باسم المئين 25، 50، و75).

Statistics

		درجة الطالب في المقرر1	درجة الطالب في المقرر2	درجة الطالب في المقرر3	عمر الطالب بالسنوات	جنس الطالب	ترتيب الفصل الدراسي للطالب	عدد أفراد أسرة الطالب	عدد الحجرات في منزل الطالب
N	Valid	35	35	35	35	35	35	35	35
	Missing	0	0	0	0	0	0	0	0
Mean		71.31	72.97	55.34	22.14	1.49	5.06	6.00	4.11
Median		75.00	75.00	53.00	22.00	1.00	5.00	5.00	4.00
Mode		75[a]	70	50[a]	22	1	4	4[a]	5
Std. Deviation		15.605	15.219	14.289	1.648	.507	1.846	3.058	1.278
Minimum		35	40	27	19	1	2	2	2
Maximum		94	96	95	25	2	8	12	6
Percentiles	25	63.00	66.00	49.00	21.00	1.00	4.00	4.00	3.00
	50	75.00	75.00	53.00	22.00	1.00	5.00	5.00	4.00
	75	84.00	85.00	60.00	23.00	2.00	7.00	9.00	5.00

a. Multiple modes exist. The smallest value is shown

شكل 23.3: المقاييس الإحصائية الاستكشافية لمتغيرات الملف "بيانات الطلبة".

ويلاحظ هنا أن النتائج أو المقاييس المحسوبة قد ظهرت بشكل فردي أو أحادي، بمعنى أن قيم المقاييس قد تم حسابها لكل متغير على حدة (في عامود مستقل)، ولا توجد نتيجة محسوبة باستخدام متغيرين أو أكثر في نفس الوقت. وبالنظر إلى هذه النتائج يمكننا استخلاص التالي:

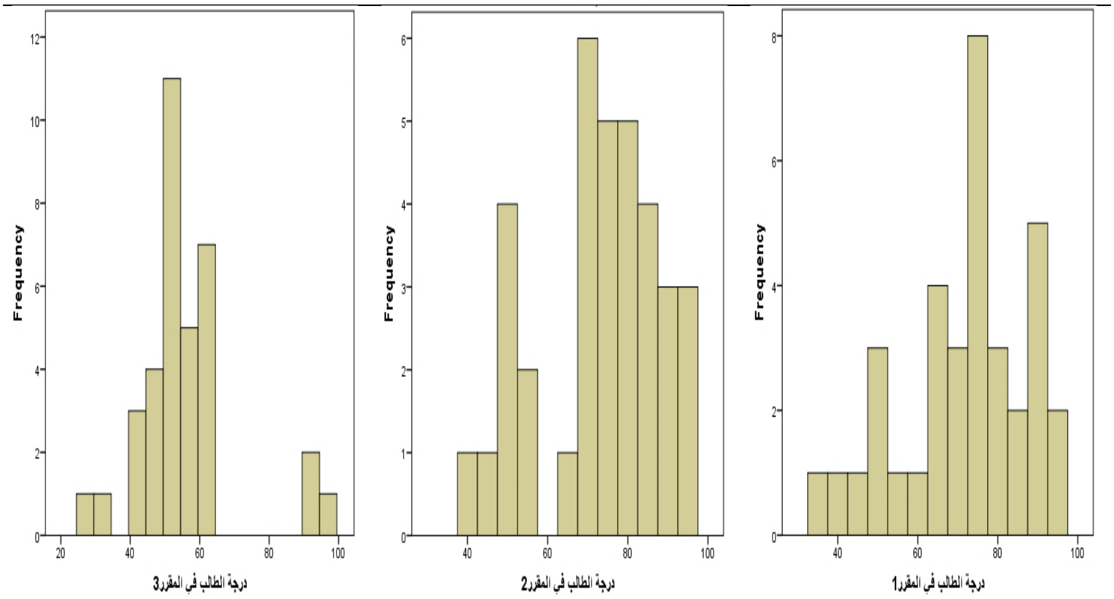
- وإذا ما نظرنا إلى الوسط الحسابي للدرجات فإننا نلاحظ أن المقررين 1 و2 لهما أوساط متقاربة تقترب من تقدير جيد جدا (71.31 و72.97)، أما المقرر الثالث فيختلف وسطه عن المقررين الأولين ويتجه نحو التقدير مقبول (55.34). وضمن عائلة الأوساط أيضا، يُلاحظ تساوي الوسيط للمقررين الأولين (75) واختلاف (ارتفاع) قيمته عن المقرر الثالث (53) بصورة كبيرة. هذه النتيجة تعطي انطباعا "مبدئيا" بأن أداء الطلاب في المقررين 1 و2 هو أفضل من أداؤهم في المقرر 3.

- الارتفاع الطفيف في قيم الوسيط للمقررين الأولين يعكس وجود التواء بسيط إلى اليسار في توزيع[1] درجات الطلاب في هذين المقررين، بمعنى أن الدرجات تتجه إلى اليمين، أي تتجه نحو الدرجات الأعلى، وهذه الملاحظة تتوافق مع قيم الأوساط للدرجات.
- إذا ما نظرنا إلى القيم الكبرى للمقررات الثلاثة والتي يُلاحظ تقاربها (94، 96، و95 على الترتيب)، فإننا، وبعد قراءة النتائج السابقة، نستنتج وجود قيم متطرفة عليا في المقرر الثالث، بمعنى أن طالب أو أكثر قد حققوا درجات عالية جدا في هذا المقرر رغم تدني مستوى الأداء لغالبية الطلاب مما يُعد من الناحية الإحصائية "تطرفا" في قيم ذلك المتغير.
- بالنظر إلى قيم الربيعات في النتيجة السابقة فإننا نلاحظ أن قيم الربيع الأول هي متقاربة في المقررين الأولين (63.5 و67.5) وهذه القيم تعني أن تقديرات 75% من الطلاب (أي الأكثرية) هي تقريبا جيد في المقررين. كما أن قيم الربيع الثالث للمقررين هي متقاربة جدا (83 و84.5) مما يدل على تقارب توزيع درجات الطلاب في هذين المقررين. أما بالنسبة للمقرر الثالث، فمن الملاحظ أن قيم الربيع الأول والثالث له متدنية كثيرا عن المقررين الأولين، وحيث أن قيمة الربيع الثالث له هي 60، (مما يعني أن 75% من درجات الطلاب في هذا المقرر هي أقل من جيد)، فهذا كله يؤكد أن أداء الطلاب في المقرر 3 كان تحت المستوى المُرضي مقارنة بالمقررين 1 و2.
- من قيم الانحراف المعياري للدرجات، نرى بأن متوسط الفروقات بين درجات الطلبة في المقررات الثلاثة هو متقارب جدا ويساوي 15 درجة تقريبا، بمعنى أن معدل الاختلاف في أداء الطلبة في أي مقرر من المقررات الثلاثة هو 15 درجة، وذلك يُعد اختلافا كبيرا بين مستوى الطلبة ضمن كل مقرر من هذه المقررات.

1.4.3 المدرج التكراري (Histogram)

إن الكثير من تلك الاستنتاجات السابقة وغيرها يمكن قراءتها "بشكل مرئي" باستخدام التمثيل البياني، وعلى رأسها المدرج التكراري الذي ستكون نتائجه متوفرة ضمن نافذة المخرجات، وسنقوم بعرض المدرجات التكرارية للمتغيرات الثلاثة الأولى في مثالنا الحالي، (ملف البيانات "بيانات الطلبة")، وهي درجات الطلبة في المقررات الثلاثة، في شكل واحد هنا، (شكل (24.3))، لتسهيل المقارنة بينها؛

[1] نذكر هنا أن مصطلح "توزيع" الدرجات يشير إلى نمط انتشار قيم المشاهدات ولا يُقصد به التوزيع الاحتمالي للمتغير.



شكل 24.3: المدرجات التكرارية للمتغيرات درجات الطلبة في مقرر 1، 2، و3 في الملف "بيانات الطلبة".

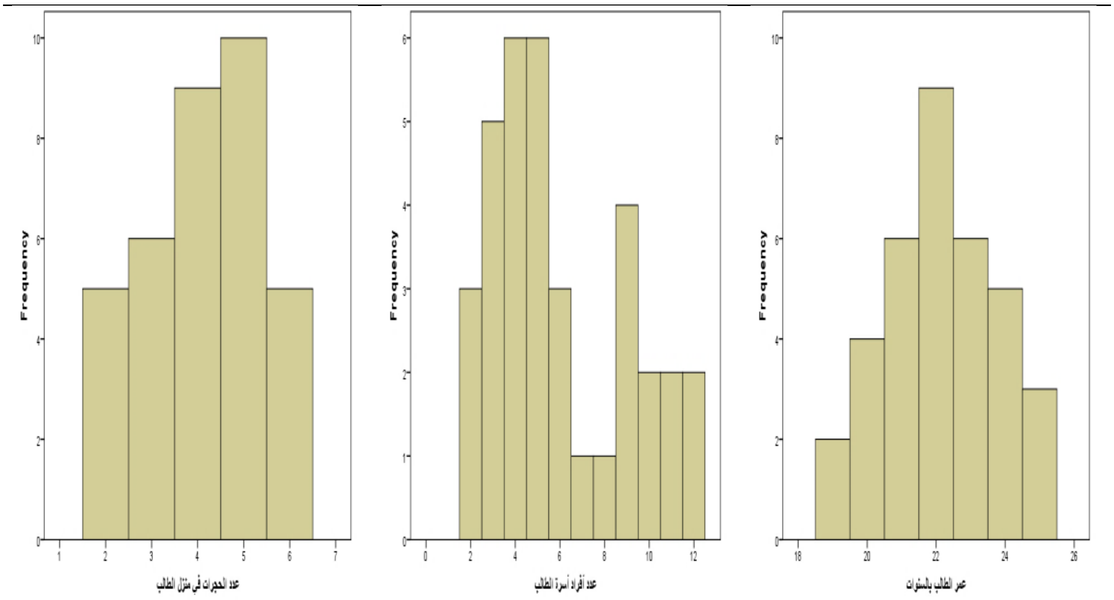
ويُلاحظ من المدرجات في الشكل (24.3) وجود التواء بسيط إلى اليسار في توزيع المقررين الأولين 1 و2، أما قيم المقرر الثالث فتُظهر التواء إلى اليمين والذي ساعد على ظهوره وجود القيم المتطرفة العليا، (العامودين الأخيرين في يمين المدرج التكراري).

أما فيما يخص باقي المتغيرات الكمية؛ وهي عمر الطالب، عدد أفراد أسرة الطالب وعدد غرف منزل الطالب، فيتم دراسة توزيعاتها بشكل منفصل أي بدون مقارنتها مع بعضها البعض، (كما كان الحال مع درجات الطلاب)، لأنها مقاسة بوحدات قياس مختلفة؛ فالعمر مُقاس بالسنة وأفراد الأسرة بعدد الأشخاص وغرف المنزل مُقاسة بعددها.

من الشكل (23.3)، يمكن ملاحظة التالي بالنسبة لمتغيرات "العمر"، "الأسرة"، و"المنزل":

- متوسط أو معدل أعمار الطلبة هو 22 سنة تقريبا، وهي أعمار منطقية تتناسب وطبيعة المرحلة الدراسية لهم. ومعدل عدد أفراد أسرة الطالب هو 6 (بوسيط يساوي 5) وهو معدل قد يُعتبر مرتفع إلى حد ما عالميا، وإن كان اعتياديا بالنسبة للأسرة العربية. أما معدل عدد الغرف في منزل الطالب في هذه العينة فيساوي 4 غرف (للوسط والوسيط) وهذا المعدل من الغرف بالنظر إلى معدل أفراد أسر الطلبة (وهو 6) يعطي انطباعا بوجود اكتظاظ في منازل هؤلاء الطلبة.
- من الانحراف المعياري للمتغيرات، نرى أن متوسط الفروقات بين أعمار الطلاب هو أقل من سنتين، كما يُلاحظ أن هنالك درجة تشتت كبيرة نوعا ما ضمن عدد أفراد أسر هؤلاء الطلاب، وأما عدد غرف منازل تلك الأسر فهي غير مختلفة كثيرا فيما بينها، حيث أن مقاييس تشتتها لها قيم منخفضة.

وبالنسبة للمدرجات التكرارية لمتغيرات العمر، عدد أفراد الأسرة، وعدد غرف المنزل، فإننا سنقوم بعرضها في الشكل التالي، (25.3)، للاختصار؛



شكل 25.3: المدرجات التكرارية لمتغيرات العمر، عدد أفراد الأسرة، وعدد غرف المنزل للطالب في الملف "بيانات الطلبة".

ومن هذه المدرجات التكرارية يمكن الوصول لنفس الملاحظات السابقة للمتغيرات الثلاثة. وننوه هنا إلى أن المدرجات التكرارية يمكن أن تستخدم لتوضيح سلوك المتغيرات من حيث مدى اقترابها من التوزيع الطبيعي كما سنوضح لاحقا.

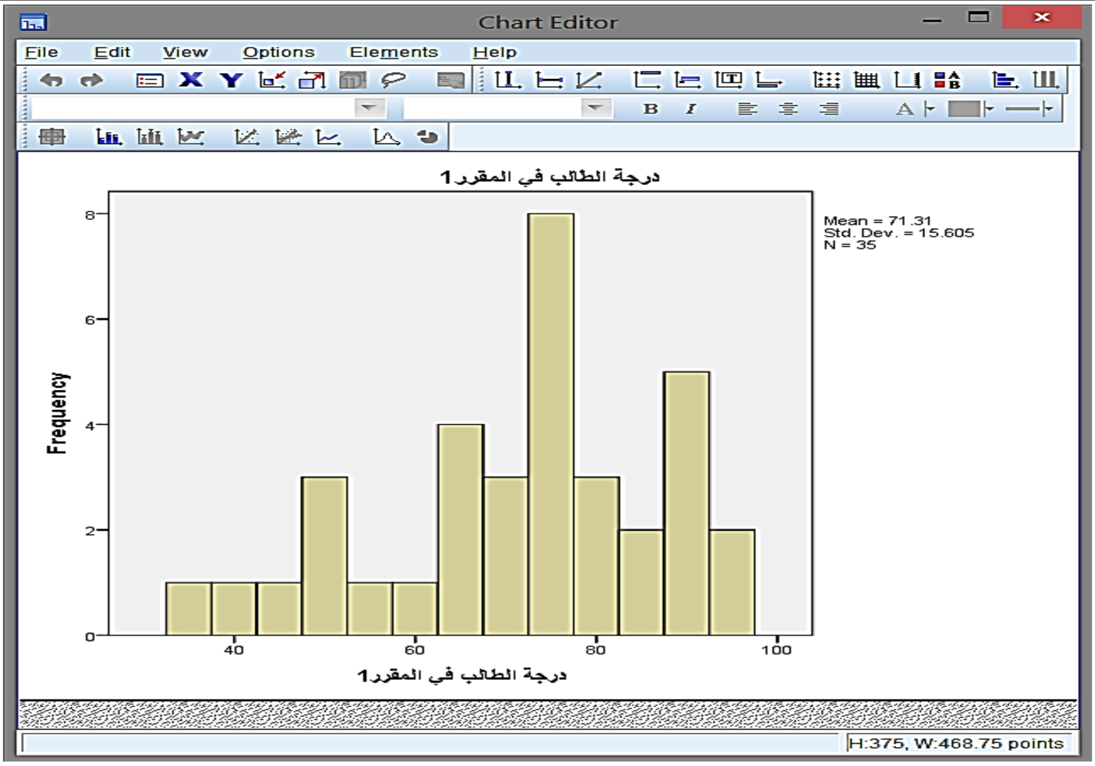
من جهة أخرى، فإنه يُلاحظ أن متغير جنس الطالب قد أعطى على سبيل المثال متوسط قدره 1.49 ووسيط يساوي 1.00 وغير ذلك من قيم للمقاييس الأخرى والسبب في ذلك يرجع إلى كون هذا المتغير وصفي وتم تعريفه ليأخذ القيمة 1 للذكور والقيمة 2 للإناث، وبالتالي فإن قيم المقاييس الإحصائية لهذا المتغير لن تكون "حقيقية" نظرا لطبيعة المتغير، بمعنى أن قيمة المتوسط 1.49 ليس لها وحدة قياس. ولذلك يفضل عادة استخدام مقاييس مثل النسبة والمنوال مع المتغيرات الوصفية. وكذلك الأمر مع المتغير الذي يمثل ترتيب الفصل الدراسي للطالب. ولذلك لن يتم عرض المدرجات التكرارية الخاصة بهاذين المتغيرين هنا.

وعموما، فإنه من الرسومات البيانية للمتغيرين يمكن القول بأن عدد الذكور هو متقارب جدا مع عدد الإناث في هذه البيانات مع وجود ارتفاع طفيف في عدد الذكور، وكذلك يُلاحظ أن معظم مسجلون في فصول دراسية في المرحلة المتوسطة من الفصل الرابع إلى الفصل السادس.

1.1.4.3 تغيير الفترات في المدرج التكراري (Adjusting the Histogram)

في بعض الحالات، قد يكون من الملائم تعديل عدد أو طول الفترات ضمن المدرج التكراري بحسب وحدة قياس المتغير، ولعمل ذلك نقوم بالآتي؛

في نفس المثال السابق، (ملف البيانات "بيانات الطلبة")، وفي نافذة المخرجات قم بالنقر المزدوج على المدرج التكراري الخاص بالمتغير "درجة الطالب في المقرر1"، فتظهر نافذة محرر الرسم (Chart Editor) كما يوضح الشكل (26.3).



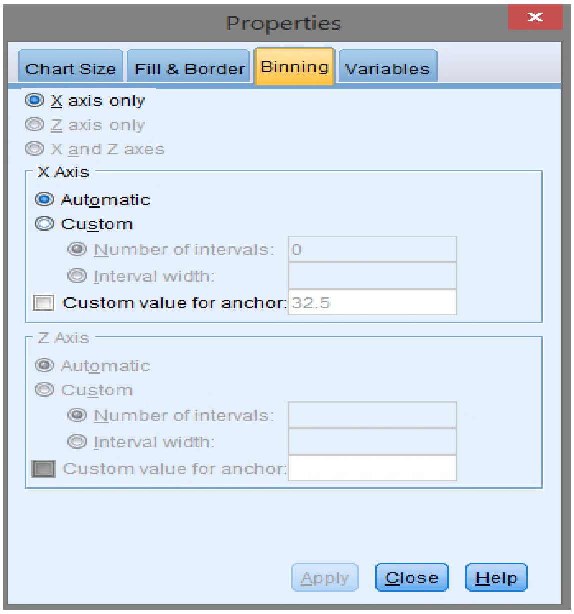
شكل 26.3: نافذة محرر الرسم للمدرج التكراري للمتغير "مقرر1".

في هذه النافذة، قم بالنقر المزدوج من جديد على المدرج التكراري بالتحديد[1] وستظهر نافذة فرعية أخرى بعنوان خصائص (Properties)، كما في الشكل (27.3). ضمن خيار تحديد المدى (Binning)، وفي الجزء الخاص بمحور X (X Axis)، اختر تعديل حسب الطلب (Custom) ثم اختر طول الفترة (Interval width)، وفي المستطيل المقابل اكتب 10 وذلك لتعديل طول الفترة إلى 10 درجات، (حيث أن المتغير يمثل درجات طلبة تتراوح ما بين 0 إلى 100 لذلك سيكون من المناسب أن يكون طول الفترة 10 درجات).

وبعد تحديد طول الفترة اضغط تطبيق (Apply) ليتم تنفيذ التعديل المطلوب، قم بعدها بإغلاق النافذة بالضغط على إغلاق (Close).

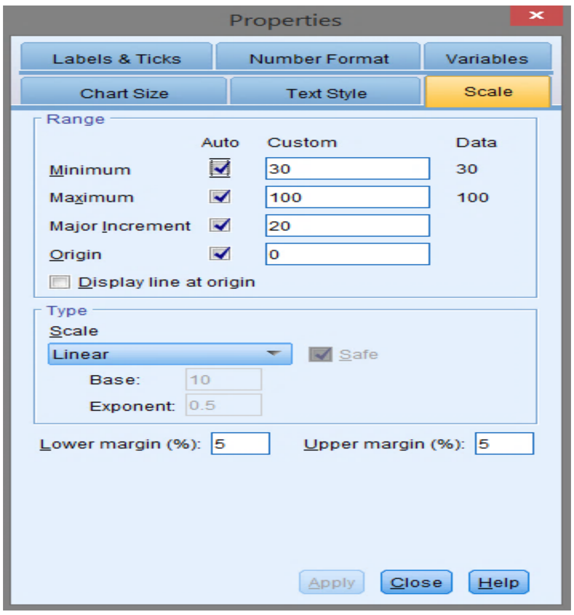
[1] نذكر هنا أن مكان النقر يحدد الجزء الذي يرغب المستخدم بتعديله، فمثلا إذا ما تم النقر على خلفية الرسم فستظهر نافذة فرعية خاصة بألوان الشكل والخلفية وغيرها، وهكذا بالنسبة للمواضع الأخرى في نافذة محرر الرسم.

ســتلاحظ الآن تغير طول فترات المدرج التكراري وكذلك تغير عدد الأعمدة في الشكل. سنقوم تاليا بتغيير مظهر قيم بداية ونهاية كل فترة (أعامدة على الرسم) بحيث تظهر كل القيم على الشكل. لتنفيذ ذلك قم ضـمن نافذة محرر الرسم السـابقة بالنقر المزدوج على أي قيمة من قيم المحور X فتظهر نافذة فرعية جديدة قم ضــمنها باختيار مقياس (Scale) فتحصــل على الخيارات المبينة في الشكل (28.3).



شكل 27.3: نافذة خصائص الفترات في محرر الرسم.

في تلك النافذة الفرعية، وفي الجزء الخاص بمدى توزيع القيم (Range)، سنقوم بتغيير مقدار الزيادة الرئيسية (Major Increment) من 20 إلى 10 درجات، ثم نضغط تطبيق فيتغير مقياس توزع القيم في محور X بحيث تظهر القيم في بداية ونهاية كل فترة، نقوم بعد ذلك بإغلاق نافذة خصائص محرر الرسم، وبذلك تكون قد انتهينا من عمل التعديلات على شــكل المدرج التكراري للمتغير المطلوب. يمكنــك الآن إغلاق نافذة محرر الرسم وإجراء تعديلات على مواضـع أخرى في الشكل نفسه أو الانتقال إلى رسم بياني آخر وإجراء أية تعديلات أخرى.



شكل 28.3: نافذة خصائص مقياس الرسم في محرر الرسم.

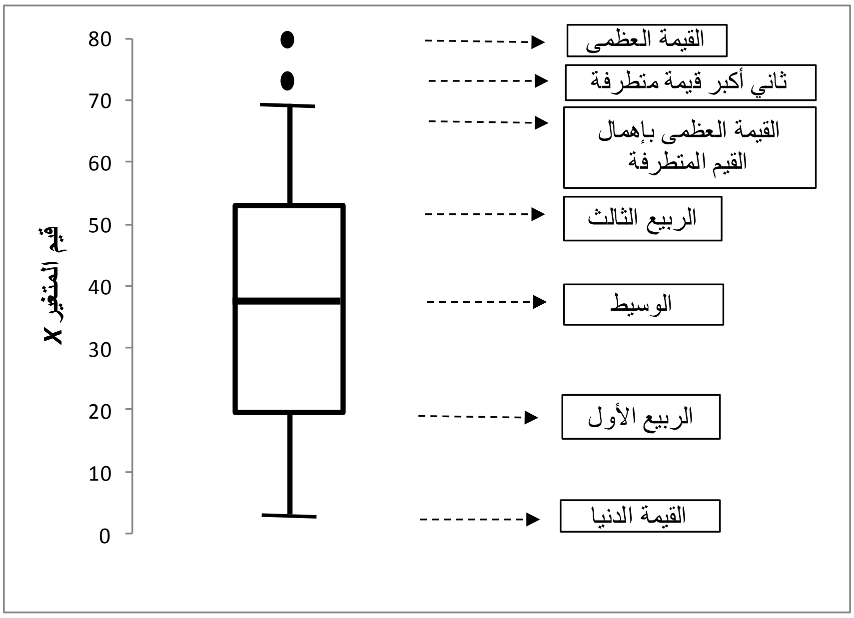
وبنفس الأســلوب يمكن تطبيق أي تعديلات على الرسومات البيانية بحسب الحاجة لها، وستلاحظ عندها بعض التغيرات في نوافذ خصــائص محرر الرسم بحسب نوع الرسم الذي تقوم بإجراء التعديل عليه.

ومن ضمن الرسومات البيانية الشهيرة في عرض البيانات ما يُعرف بشكل الصندوق والذي يقدم تقريبا نفس المعلومات التي يقدمها المدرج التكراري مع بعض الإضافات الهامة، ويمكننا تقديم تعريف مبسط لهذا التمثيل البياني الهام بالصورة التالية؛

2.4.3 شكل الصندوق (The Boxplot)

شكل الصندوق هو ببساطة رسم بياني يتم من خلاله توضيح خمسة مقاييس هامة تمثل بحد ذاتها ملخصا، قد يكون بمفرده كافيا في بعض الأحيان، لتوضيح سلوك البيانات. وهذا الرسم يمكن استخدامه لوصف مجموعة واحدة (متغير واحدة) من البيانات أو للمقارنة بين متغيرين أو أكثر، إذ يمكن من خلاله مراقبة تشتت توزيع البيانات عن مركزها وكذلك ملاحظة مدى تماثل توزيعها، وأيضا من خلال شكل الصندوق يمكننا التعرف على القيم المتطرفة بصورة واضحة جدا كما سنرى.

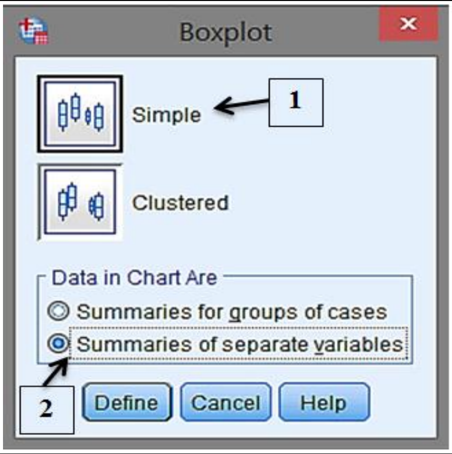
إن المقاييس الخمسة التي يعرضها الصندوق، والتي تسمى أيضا **بملخص الأرقام الخمسة** (Five-Numbers Summary)، هي القيمتين الدنيا والكبرى، الربيع الأول (Q_1)، الربيع الثالث (Q_3)، والوسيط ($\tilde{X}$). ويكون الشكل التقليدي لرسم الصندوق على الصورة التالية، (شكل (29.3))؛



شكل 29.3: شكل بياني افتراضي لرسم الصندوق موضحا عليه المقاييس الإحصائية المرتبطة به.

ويلاحظ من الشكل (29.3) أنه إذا لم توجد قيم متطرفة في توزيع البيانات، فإن القيمة العظمى ستكون هي القيمة عند الخط الذي يمثل الحد الأعلى في الصندوق، وكذلك الحال بالنسبة للقيمة الدنيا في البيانات. ولتوضيح كيفية استخدام شكل الصندوق لتمثيل توزيع المتغير بيانيا في برنامج SPSS، سنقوم باستخدامه لتمثيل المتغيرات الثلاثة الأولى من ملف البيانات "بيانات الطلبة"؛

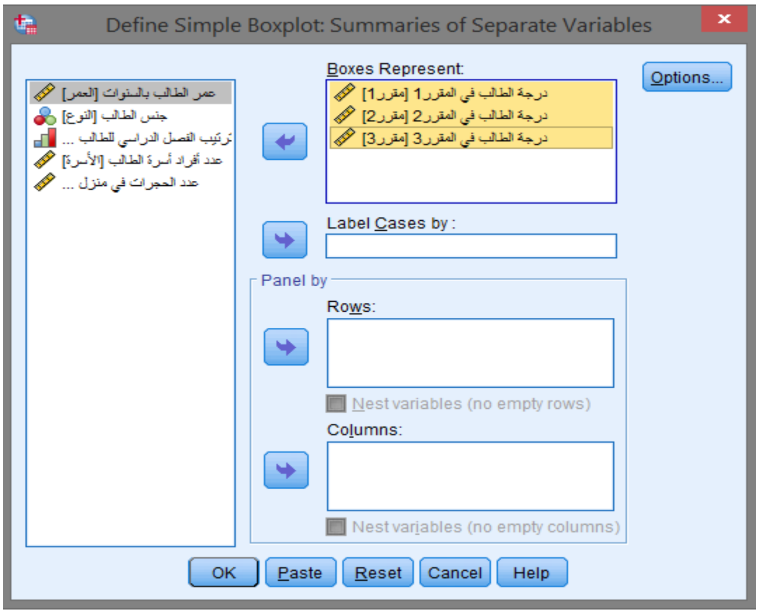
في الشريط العلوي لبرنامج SPSS قم باختيار Graphs>Legacy Dialogs>Boxplot فتظهر النافذة التالية، (شكل (30.3))؛



شكل 30.3: نافذة خيارات شكل الصندوق.

سنقوم هنا باختيار تنفيذ شكل الصندوق البسيط (Simple) والمشار إليه بالسهم رقم (1) في الشكل، (وسنتطرق لشكل الصندوق باستخدام التصنيف تاليا)، ونختار تمثيل عدة متغيرات في آن واحد[1] كما يشير السهم رقم (2) في الشكل. ثم نضغط أمر تعريف (Define)، فتظهر نافذة جديدة كما يوضح الشكل (31.3).

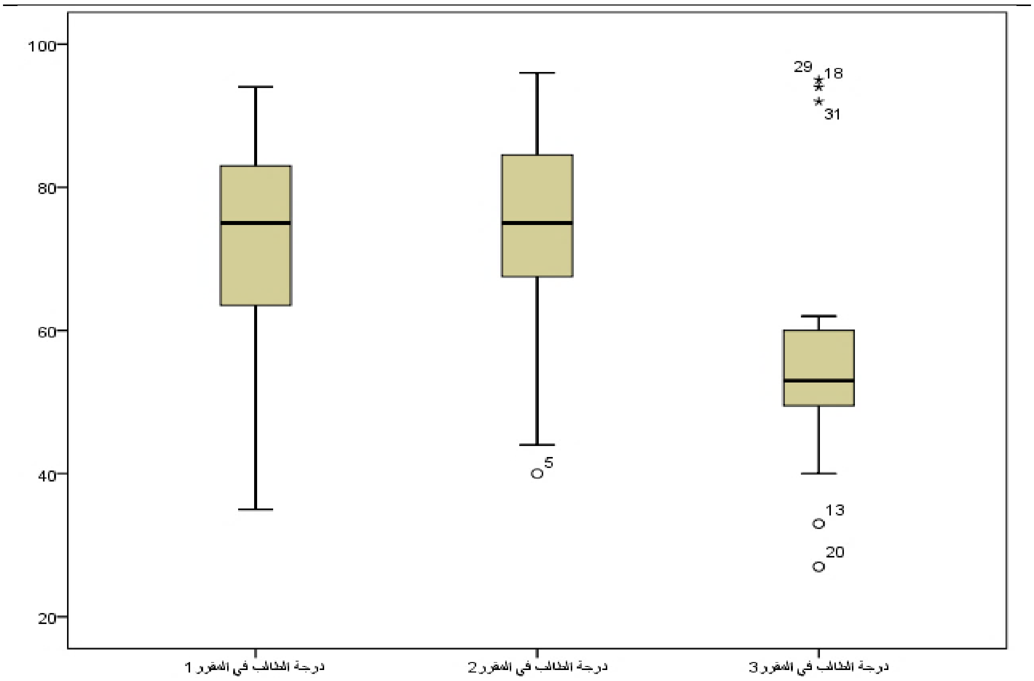
قم في تلك النافذة باختيار المتغيرات الثلاثة الأولى، (وهي درجة الطالب في المقرر 1 والمقرر 2 والمقرر 3)، ونقلها إلى الخانة العليا في يمين النافذة والمعنونة: الصناديق تمثل (Boxes Represents) كما هو مبين في الشكل، ثم اضغط موافق (OK).



شكل 31.3: نافذة تحديد المتغيرات لشكل الصندوق، (لعدة متغيرات آنيا).

فيظهر الشكل المطلوب في نافذة المخرجات، كما يوضح الشكل (32.3). ونلاحظ من الشكل ما يلي:

[1] نشير هنا إلى أن اختيار التمثيل البياني لعدة متغيرات في آن واحد يتم استخدامه فقط عندما تكون المتغيرات لها نفس وحدة القياس وليست من وحدات قياس مختلفة.



شكل 32.3: شكل الصندوق للمتغيرات درجات الطلبة في مقرر 1، 2، و3 في الملف "بيانات الطلبة".

1. اقتراب الصندوق، والذي يمثل كتلة البيانات، في المتغيرين أو المقررين الأولين 1 و2 إلى الجانب الأعلى في الرسم مما يُظهر التواء إلى اليسار في توزيع البيانات، وهذه دلالة على ارتفاع المستوى الدراسي للطلاب في هذين المقررين. إلا أن ما نراه في المتغير الثالث (المقرر3) ليس التواء بالمعنى الصحيح، لأن كل المشاهدات في هذا المتغير، باستثناء النقاط (الدرجات) الثلاثة المتطرفة (حيث تمثل القيم على الرسم ترتيب القيمة المتطرفة) في أعلى الرسم، تنحصر في الجانب الأسفل.

2. أغلبية درجات الطلاب في المقررين 1 و2 تنحصر في الفترة (65 إلى 85) تقريبا، أما أغلبية الدرجات في المقرر3 فتنحصر في الفترة (50 إلى 60)، مما يدل على انخفاض مستوى الطلاب في المقرر الثالث، كما ذكرنا سابقا.

3. عدم ظهور قيم متطرفة في المقرر الأول، وظهور قيمة واحدة متطرفة (هي الدرجة 40) في المقرر الثاني، وكذلك ظهور قيم متطرفة دنيا (هما الدرجتان 27 و33) في المقرر الثالث والتي قد لا تظهر في رسومات أخرى بوضوح، إضافة للقيم المتطرفة الثلاثة الكبرى الظاهرة.

4. انتشار البيانات (ممثلا بطول الصندوق) والذي يمثل تشتت درجات الطلاب يبدو أكبر في المقررين 1 و2 مما هو عليه في المقرر3، إلا أن ذلك لم يكن جليا واضحا في النتائج والرسومات السابقة، ولا حتى بحساب الانحراف المعياري للمقررات الثلاثة؛ (من الشكل (23.3)، الانحرافات المعيارية للمقررات الثلاثة على الترتيب هي 15.605، 15.219، و14.289 درجة)، والتي تبدو متقاربة، وهذه في الواقع إحدى مزايا التمثيل الجيد لتوزيع البيانات في

رسم الصندوق. وننوه هنا إلى أن اقتراب قيمة الانحراف المعياري للمقرر الثالث من قيم المقررين الأولين سببه وجود القيم المتطرفة العليا التي أدت لرفع قيمته.

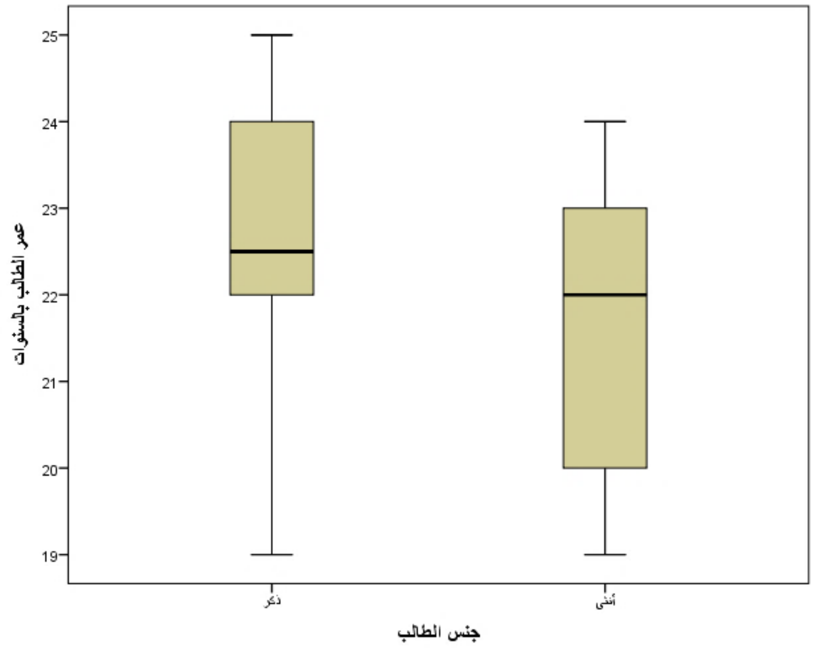
من جديد، يمكن استخدام التمثيل البياني لشكل الصندوق بطريقة تُظهر توزيع المتغير الأساسي مصنفا بحسب متغير (أو متغيرين)، فمثلا يمكننا باستخدام بيانات المثال الحالي تنفيذ شكل الصندوق لمتغير عمر الطالب بحسب النوع؛ ذكر أو أنثى.

لعمل ذلك، قم في الشريط العلوي لبرنامج SPSS باختيار Graphs>Legacy Dialogs>Boxplot وفي النافذة التي تظهر في الشكل (30.3) قم باختيار تنفيذ شكل الصندوق البسيط (Simple) والمشار إليه بالسهم رقم (1)، ثم قم باختيار الخيار الأول وهو خيار تمثيل متغير واحد (Summaries of groups of cases). ثم نضغط أمر تعريف (Define)، فتظهر نافذة جديدة كما هو في الشكل (33.3).



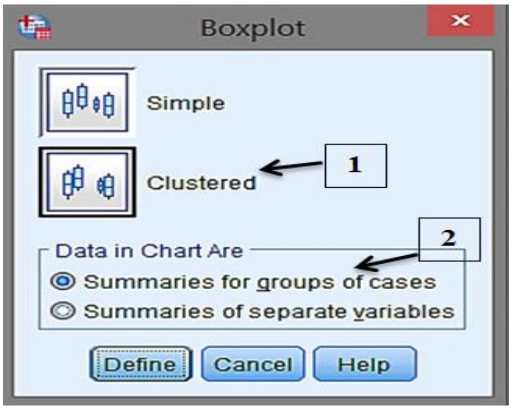
شكل 33.3: نافذة تحديد المتغيرات لشكل الصندوق، (لمتغير واحد).

قم في تلك النافذة باختيار المتغير عمر الطالب في الخانة الأولى العليا (Variable)، (كما يوضح الشكل)، ثم قم باختيار متغير التصنيف أو التقسيم وهو متغير جنس الطالب في الخانة الثانية (Category Axis)، ثم اضغط موافق فيظهر شكل الصندوق المصنف بحسب جنس الطالب في نافذة المخرجات، كما يوضح الشكل (34.3). ونلاحظ من الشكل بوضوح أن الطلبة الذكور في العينة هم في الغالب أكبر سنا من الطالبات.



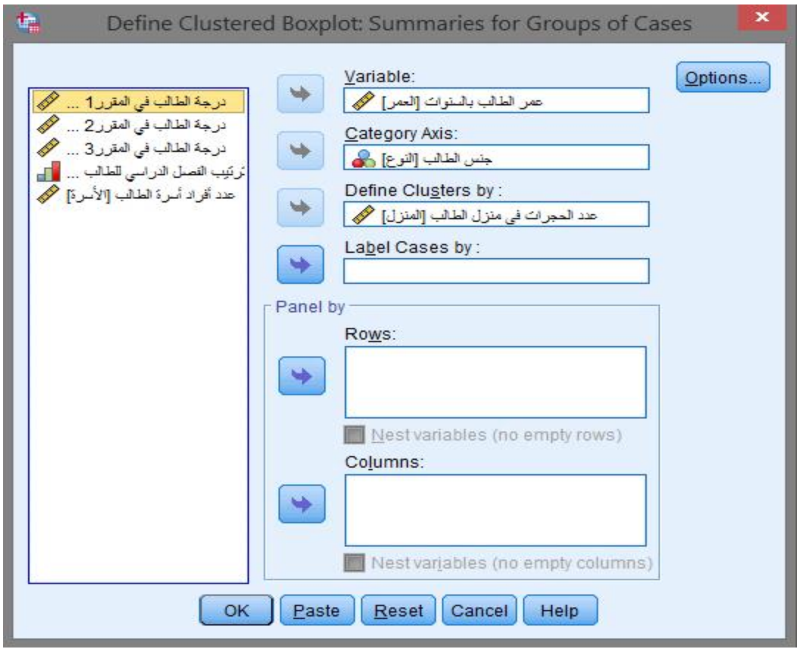
شكل 34.3: شكل الصندوق المصنف لعمر الطالب بحسب النوع للمتغيرات في الملف "بيانات الطلبة".

ويمكن أيضا، في نفس السياق، استخدام رسم الصندوق لتمثيل متغير ما اعتمادا على تصنيفه باستخدام متغيرين وليس متغير واحد فقط، (كما شاهدنا في الشكل السابق (34.3)). فمثلا يمكننا في نفس المثال الحالي استخدام شكل الصندوق لتمثيل متغير عمر الطالب بناء على متغير الجنس ومتغير عدد حجرات منزله. لتنفيذ ذلك، قم في الشريط العلوي لبرنامج SPSS باختيار Graphs>Legacy Dialogs>Boxplot وفي النافذة التي تظهر في الشكل (35.3) قم باختيار تنفيذ شكل الصندوق المصنف (Clustered) والمشار إليه بالسهم رقم (1)، ثم قم باختيار الخيار الأول والمشار إليه بالسهم رقم (2) وهو خيار تمثيل متغير واحد. ثم اضغط أمر تعريف (Define)، فتظهر نافذة جديدة كما هو في الشكل (36.3).



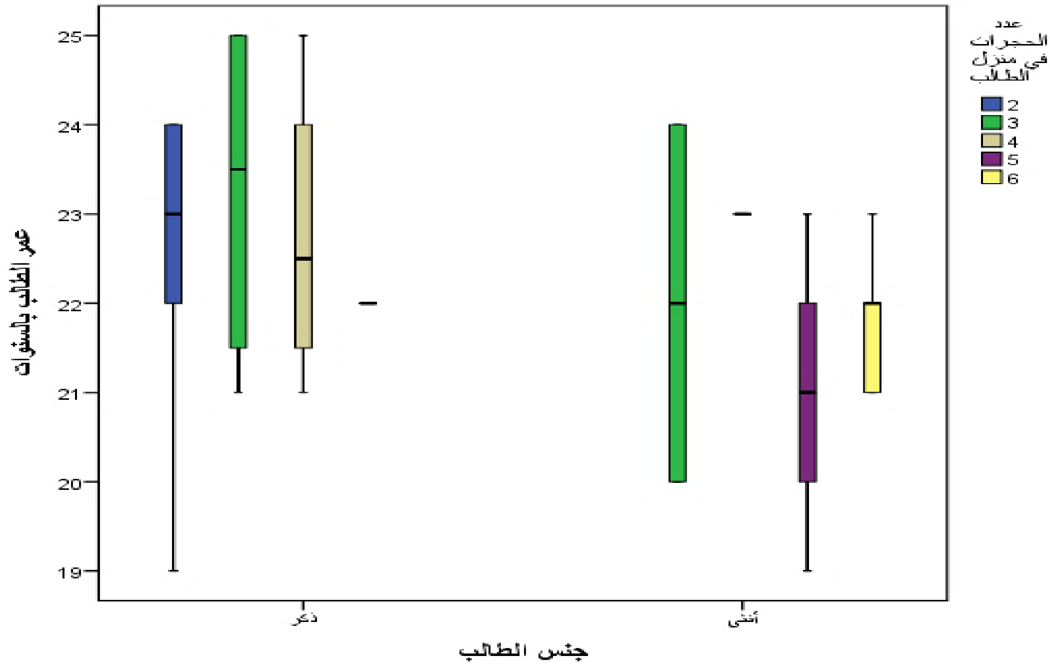
شكل 35.3: نافذة خيارات شكل الصندوق مع اختيار رسم الصندوق المصنف.

قم في تلك النافذة باختيار المتغير عمر الطالب في الخانة الأولى العليا (Variable)، ثم قم باختيار متغير التصنيف الأول وهو متغير جنس الطالب في الخانة الثانية (Category Axis)، ثم اختر متغير التصنيف الثاني وهو عدد الحجرات في منزل الطالب في الخانة الثالثة (Define Clusters by)، ثم اضغط موافق فيظهر شكل الصندوق المصنف بحسب كلا من جنس الطالب (في المحور الأفقي) وبحسب عدد حجرات المنزل (مقسما للذكور وللإناث كل على حده) في نافذة المخرجات، كما يوضح الشكل (37.3).



شكل 36.3: نافذة تحديد المتغيرات لشكل الصندوق، (لمتغير واحد مصنف بحسب متغيرين).

ونلاحظ من الشكل ارتفاع أعمار الطلبة الذكور في العينة في العموم، كما أنه يمكن القول بأن الطلبة الذكور في الغالب يقطنون في منازل أعداد حجراتها تتراوح ما بين حجرتين وأربعة حجرات، وأما الطالبات فيقطن في منازل عدد حجراتها هو ثلاثة أو خمسة أو ستة حجرات على الأكثر.



شكل 37.3: شكل الصندوق لعمر الطالب مصنفا بحسب النوع وعدد حجرات المنزل للطالب للمتغيرات في الملف "بيانات الطلبة".

3.4.3 الرسم الخطي أو مخطط الزمن (Line or Time Chart)

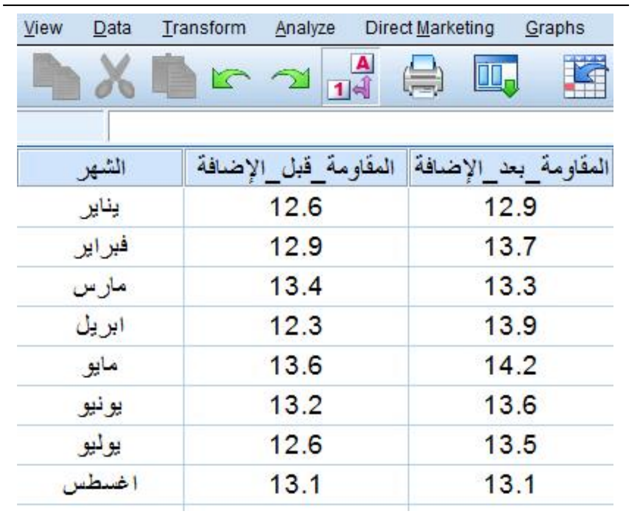
وهو نوع من الرسومات البيانية التي تهدف إلى توضيح التغير في مفردات البيانات خلال فترة زمنية معينة (سنوات، أشهر ...، أو حتى مواسم). ويستخدم هذا النوع من الرسومات في الكثير من المجالات الإنتاجية أبرزها عرض التقارير الاقتصادية والصناعية. والتمثيل البياني لمخطط الزمن يتم برصد نقطة لكل قيمة من قيم المتغير مقابل كل وحدة زمنية ثم توصيل هذه النقاط بخطوط مستقيمة. ولتوضيح كيفية تنفيذ هذا النوع من الرسومات، لنأخذ المثال التالي؛

أثناء إجراء دراسة تتعلق باختبار قوة تحمل خليط من المعادن المستخدمة في بناء جدران المنشئات العسكرية، تم في البداية عمل نموذج أولي وتم اختباره 8 مرات خلال ثمانية أشهر متتالية مع زيادة درجة الحرارة المحيطة بالتدريج. ثم تم بعد ذلك إضافة معدن جديد للخليط وإعادة الاختبار تحت نفس درجات الحرارة المستخدمة مع الخليط الأول فكانت النتائج، (والتي تمثل مقاومة الخليط بالكيلوجرام)، كما هو موضح في الجدول المرفق (جدول (5.3))؛

جدول 5.3: نتائج مقاومة خليط المعادن قبل وبعد إضافة المعدن الجديد خلال فترة ثمانية أشهر.

الشهر	يناير	فبراير	مارس	ابريل	مايو	يونيو	يوليو	أغسطس
(المقاومة قبل الإضافة)	12.6	12.9	13.4	12.3	13.6	13.2	12.6	13.1
(المقاومة بعد الإضافة)	12.9	13.7	13.3	13.9	14.2	13.6	13.5	13.1

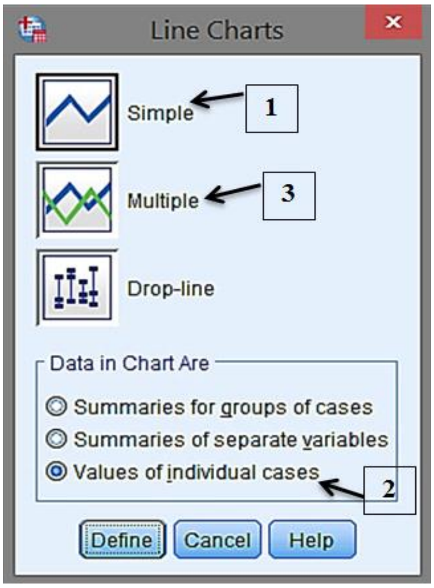
وسنقوم باستخدام مخطط الزمن لتمثيل التغير في مقاومة خليط المعدن قبل الإضافة خلال الأشهر المتلاحقة. وستكون الخطوة الأولى هي إدخال البيانات (في جدول (5.3)) في ملف بيانات جديد في SPSS، وليكن باسم "مقاومة المعدن" كما يظهر في الشكل (38.3).



الشهر	المقاومة_قبل_الإضافة	المقاومة_بعد_الإضافة
يناير	12.6	12.9
فبراير	12.9	13.7
مارس	13.4	13.3
ابريل	12.3	13.9
مايو	13.6	14.2
يونيو	13.2	13.6
يوليو	12.6	13.5
اغسطس	13.1	13.1

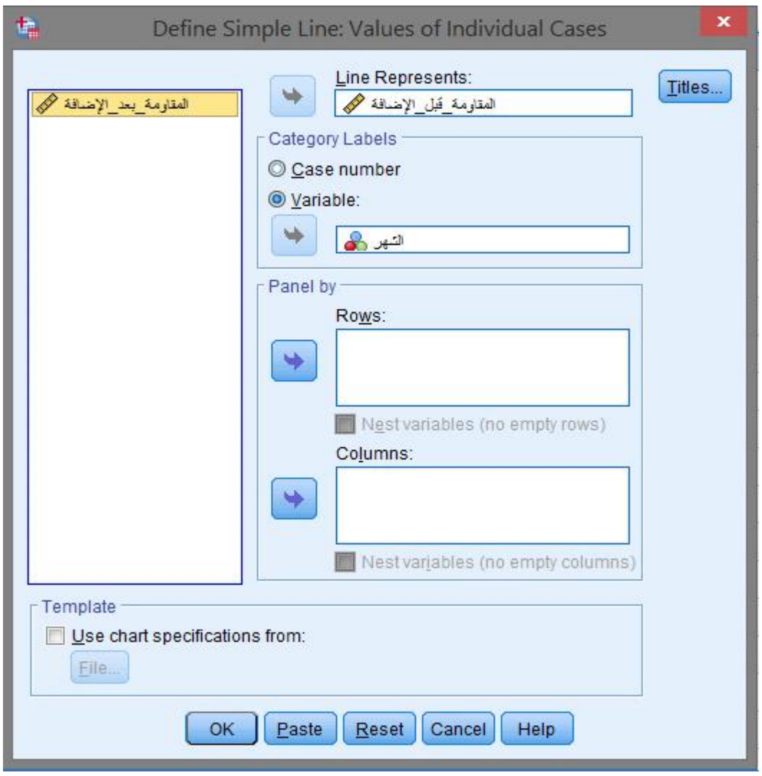
شكل 38.3: ملف بيانات "مقاومة المعدن".

بعد إدخال البيانات، قم في شريط الأدوات العلوي باختيار Graphs>Legacy Dialogs>Line، فتظهر نافذة خيارات الرسم الخطي كما هو في الشكل (39.3). قم باختيار تنفيذ شكل الرسم الخطي البسيط (Simple) والمشار إليه بالسهم رقم (1)، ثم قم بتحديد الخيار الخاص بتمثيل قيم المشاهدات (Values of individual cases) والمشار إليه بالسهم رقم (2).



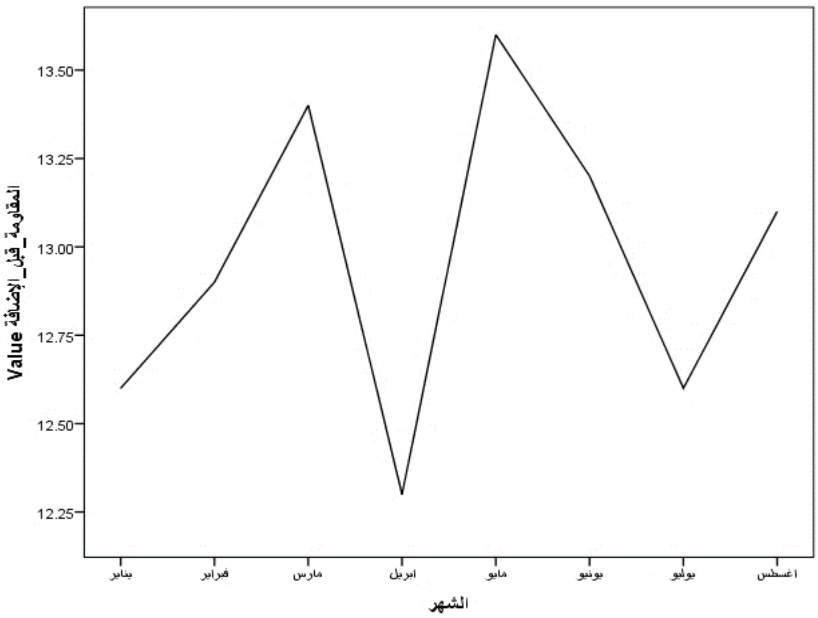
شكل 39.3: نافذة خيارات الرسم الخطي.

اضغط بعد ذلك على تعريف (Define) فتظهر نافذة تعريف المتغيرات الخاصة بالرسم المطلوب، (الشكل (40.3))، في تلك النافذة سنقوم باختيار المتغير (المقاومة_قبل_الإضافة) في المربع الأول في الأعلى: الخط يمثل (Line Represents)، ومن أسفل ذلك المربع نختار خيار (Category Labels) متغير (Variable)، ثم نختار متغير "الشهر" في الشريط أدناه فنحصل على الشكل البياني المطلوب، (شكل (41.3)).



شكل 40.3: نافذة تحديد المتغيرات للرسم الخطي لمتغير واحد.

من الرسم الخطي أو مخطط الزمن في الشكل (41.3)، نلاحظ وجود تذبذب في ارتفاع وانخفاض مقاومة المعدن قبل الإضافة خلال الأشهر المتتالية، فنلاحظ ارتفاع مقاومة المعدن من تجارب شهر يناير حتى شهر مارس ثم انخفاض مقاومة المعدن في شهر ابريل ثم ارتفاعها في مايو وهكذا.

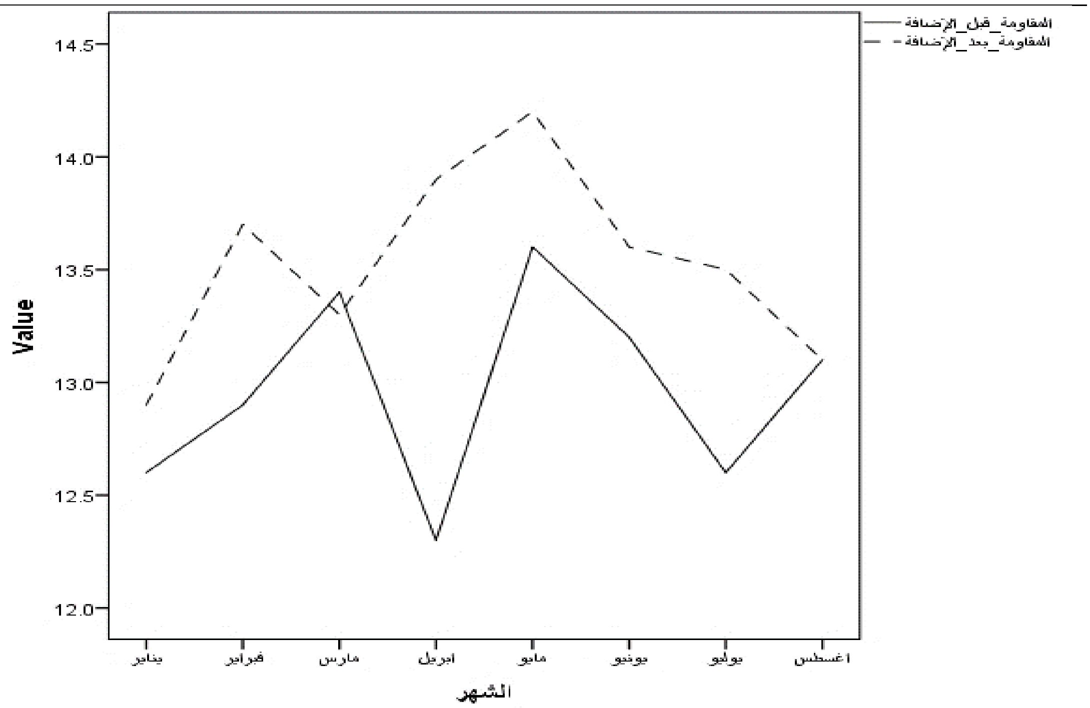


شكل 41.3: الرسم الخطي لمتغير المقاومة_قبل_الإضافة في ملف "مقاومة المعدن".

ويمكن أيضا استخدام الرسم الخطي لمقارنة التغير عبر الزمن بين متغيرين (أو أكثر)، ففي المثال الحالي، يمكننا مقارنة مقاومة المعدن قبل وبعد إضافة المعدن الجديد. ولتنفيذ ذلك، نقوم في شريط الأدوات العلوي باختيار Graphs>Legacy Dialogs>Line، فتظهر نافذة خيارات الرسم الخطي كما هو في الشكل (39.3) السابق.

نقوم باختيار تنفيذ شكل الرسم الخطي المتعدد (Multiple) والمشار إليه بالسهم رقم (3)، ثم نقوم بتحديد الخيار الخاص بتمثيل قيم المشاهدات والمشار إليه بالسهم رقم (2). نضغط بعد ذلك على تعريف (Define) فتظهر نافذة تعريف المتغيرات الخاصة بالرسم المطلوب كما هو موضح في الشكل (40.3) السابق.

في تلك النافذة نقوم باختيار كلا من المتغيرين (المقاومة_قبل_الإضافة) و(المقاومة_بعد_الإضافة) في الشريط الأول في الأعلى (Line Represents)، ومن أسفل ذلك الشريط (Category Labels) نختار خيار متغير (Variable)، ثم نختار متغير "الشهر" في الشريط أدناه فنحصل على الرسم الخطي المتعدد، (شكل (42.3)). ويُلاحظ من الرسم أن إضافة المعدن الجديد قد أدى إلى زيادة مقاومة الخليط بصورة عامة عبر الأشهر الثمانية.



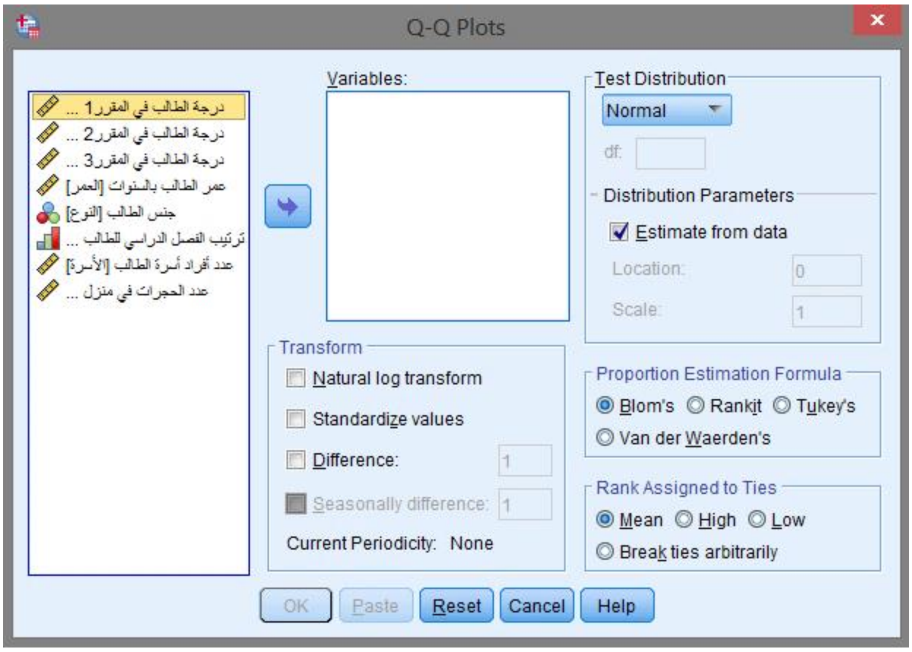
شكل 42.3: الرسم الخطي لمتغيري المقاومة_قبل_الإضافة و المقاومة_قبل_الإضافة في ملف "مقاومة المعدن".

4.4.3 طبيعية توزيع المتغيرات (Normality of Variables)

تعرفنا فيما سبق على استخدام المدرج التكراري وشكل الصندوق لاستكشاف توزيع المتغيرات الكمية، ومن ضمن استخدامات هذين التمثيلين التعرف على التوزيع الاحتمالي للمتغير ما إذا كان طبيعيا أم لا. إلا أن كثير من المتخصصين يفضلون استخدام أسلوب آخر ضمن طرق الرسم البياني[1] للتحقق من توزع المتغير طبيعيا وهو رسم Q-Q وP-P، (Q-Q and P-P Normal Probability Plot) حيث يمثل الأول رسم الربيعات والثاني رسم المئينات للمتغير ومطابقتها بقريناتها في التوزيع الطبيعي. ولتوضيح الصورة لنأخذ المثال التالي؛

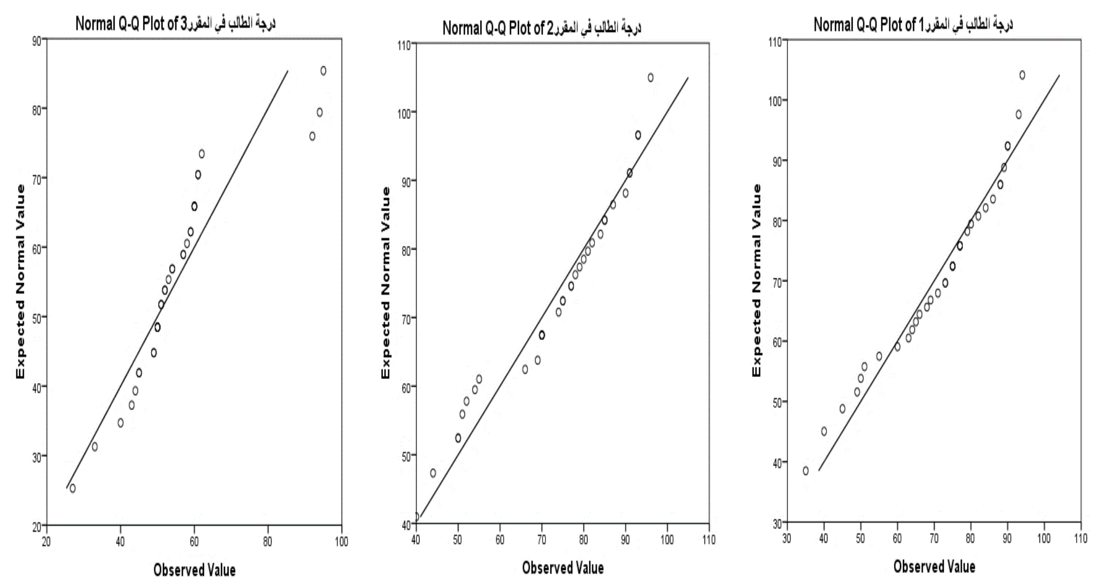
في ملف البيانات "بيانات الطلبة"، سنقوم بتطبيق رسم Q-Q للمقارنة بين مدى توزع المتغيرات الثلاثة طبيعيا. من شريط الأدوات العلوي قم باختيار Analyze>Descriptive Statistics>Q-Q Plots، فتظهر عندها نافذة رسم Q-Q كما هو مبين في الشكل (43.3). في القسم العلوي الأيمن من النافذة (Test Distribution) يمكنك ملاحظة وجود الخيار الخاص باختيار التوزيع الاحتمالي الذي ترغب برسم Q-Q له، وفي مثالنا الحالي نرغب برسم Q-Q لبعض المتغيرات ومقارنتها بالتوزيع الطبيعي، والذي يمثل الخيار الافتراضي للرسم.

[1] يمكن الإجماع على أن أفضل الأساليب الإحصائية للتأكد من التوزيع الاحتمالي للمتغيرات، ومن ضمنها التوزيع الطبيعي، هي اختبارات الفروض. وسوف نتطرق للاختبارات الخاصة بالتوزيع الطبيعي في الفقرة القادمة.



شكل 43.3: نافذة رسم Q-Q لملف البيانات "بيانات الطلبة".

سنقوم باختيار المتغيرات الثلاثة الأولى، وهي درجات الطلبة في ثلاثة مقررات، ونقلها إلى خانة المتغيرات، وبعدها نضغط موافق فنحصل على عدة نتائج من أهمها المتوسط والانحراف المعياري للمتغيرات إضافة لرسم Q-Q لكل متغير. وسنعرض في الشكل (44.3) التالي الرسومات للمتغيرات جنبا إلى جنب لتسهيل المقارنة.



شكل 44.3: رسم Q-Q لدرجات الطلبة في المقررات الثلاثة في الملف "بيانات الطلبة".

في رسم Q-Q أو P-P يتم التدقيق في مدى اقتراب النقاط، (والتي تمثل المشاهدات للمتغير)، من الخط القطري المرسوم، (والذي يمثل توزيع النقاط الافتراضي للتوزيع الطبيعي)، وبالتالي كلما اقتربت نقاط المتغير من الخط القطري كلما دل ذلك على اقتراب توزيع المتغير من التوزيع الطبيعي. ومن الشكل (44.3) يمكننا القول بأن درجات

الطلبة في المقرر 1 هي الأكثر اقترابا من التوزيع الطبيعي مقارنة بدرجات الطلبة في المقررين الآخرين، ونلاحظ أيضا ابتعاد المتغير الثالث، (درجات الطلبة في المقرر 3)، عن التوزيع الطبيعي بصورة كبيرة. إلا أنه، وكما ذكرنا سابقا، يكون من المفضل إحصائيا اختبار توزيع المتغيرات طبيعيا باستخدام اختبارات الفروض وعدم الاعتماد كليا على الرسم البياني كما سنرى في الفقرة التالية.

5.3 استخدام أسلوب استكشاف البيانات الشامل في SPSS (Using the Explore command in SPSS)

لاحظنا في الأمثلة السابقة كيفية استخدام المقاييس الإحصائية الوصفية والرسومات البيانية عن طريق اختيار أوامر التكرارات (Frequencies) والإحصاءات الوصفية (Descriptives) الموجودة تحت قائمة (Analyze>Descriptive Statistics) في شريط الأوامر العلوي في البرنامج. والآن سنعرّف القارئ بكيفية استكشاف توزيع أو سلوك متغيرات الدراسة بصورة أعم وأشمل عن طريق استخدام أمر **استكشاف** (Explore) الموجود ضمن نفس القائمة السابقة.

وكمثال تطبيقي لاستكشاف البيانات، سنستخدم الجدول التالي، جدول (6.3)، والذي يضم بيانات تمثل عينة مكونة من 90 طالبا معظمهم من الطلبة القدامى، (الذين تأخر تخرجهم عن المدة الزمنية المعتادة، وهي ثمانية فصول دراسية تقريبا)، في كلية العلوم بجامعة بنغازي مقسمة بحسب التخصص أو القسم العلمي، (رياضيات (MA)، إحصاء (ST)، وفيزياء (PH))، وتشمل معدل الطالب، (على مقياس من 0.00 إلى 4.00 درجة)، جنس الطالب، وعدد الفصول الدراسية التي أنجزها الطالب.

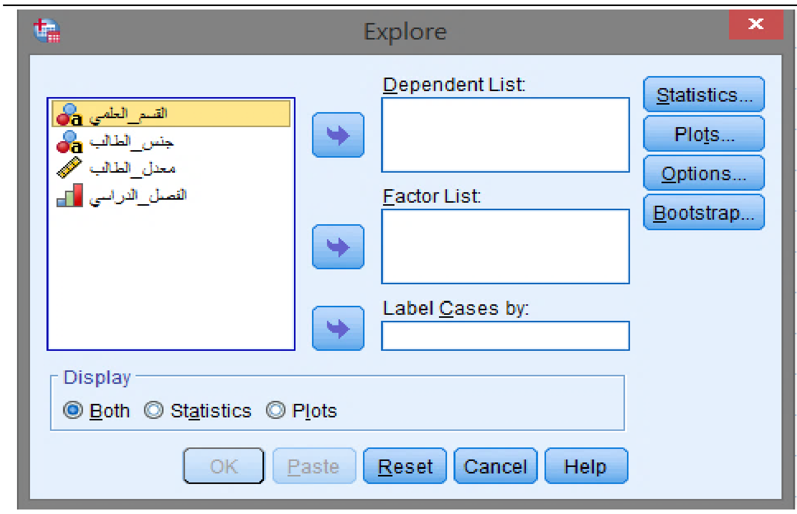
جدول 6.3: معدلات عينة من 90 طالبا من طلبة كلية العلوم في جامعة بنغازي.

م	القسم العلمي	جنس الطالب	معدل الطالب	الفصل الدراسي	م	القسم العلمي	جنس الطالب	معدل الطالب	الفصل الدراسي	م	القسم العلمي	جنس الطالب	معدل الطالب	الفصل الدراسي
1	MA	M	1.96	20	31	ST	F	1.60	12	61	PH	F	1.75	23
2	MA	M	1.57	18	32	ST	M	1.51	14	62	PH	F	1.67	19
3	MA	F	1.48	16	33	ST	F	1.98	14	63	PH	F	1.58	21
4	MA	F	1.58	21	34	ST	M	2.18	18	64	PH	F	1.41	21
5	MA	F	1.80	18	35	ST	F	1.64	18	65	PH	F	1.61	20
6	MA	F	1.26	15	36	ST	F	1.68	16	66	PH	M	1.53	20
7	MA	F	1.69	16	37	ST	F	1.90	17	67	PH	F	1.68	17
8	MA	M	1.26	14	38	ST	F	1.54	18	68	PH	F	1.46	19
9	MA	F	1.60	12	39	ST	M	1.56	13	69	PH	F	1.26	20
10	MA	F	1.89	13	40	ST	F	1.56	13	70	PH	F	1.62	19

(تابع) جدول 6.3: معدلات عينة من 90 طالبا من طلبة كلية العلوم في جامعة بنغازي.

11	MA	F	1.48	12	41	ST	F	1.88	13	71	PH	F	1.57	18
12	MA	F	2.25	11	42	ST	F	1.16	13	72	PH	F	2.09	13
13	MA	M	1.63	11	43	ST	F	1.74	12	73	PH	F	1.57	19
14	MA	F	1.50	11	44	ST	F	1.15	12	74	PH	F	1.45	19
15	MA	F	1.78	11	45	ST	F	1.25	13	75	PH	F	1.67	18
16	MA	F	1.84	11	46	ST	M	1.55	13	76	PH	F	1.78	18
17	MA	M	2.40	11	47	ST	M	1.45	13	77	PH	F	1.68	18
18	MA	M	2.09	11	48	ST	M	1.75	13	78	PH	F	1.53	17
19	MA	F	1.74	11	49	ST	F	2.01	13	79	PH	M	1.66	16
20	MA	F	1.52	9	50	ST	M	1.40	13	80	PH	F	1.66	16
21	MA	M	1.79	9	51	ST	M	1.63	13	81	PH	F	1.84	12
22	MA	F	3.15	8	52	ST	F	1.87	12	82	PH	F	1.79	16
23	MA	F	1.95	2	53	ST	M	1.78	5	83	PH	F	1.63	16
24	MA	F	3.39	7	54	ST	M	1.58	12	84	PH	F	1.73	16
25	MA	F	1.87	7	55	ST	F	1.88	11	85	PH	F	1.60	16
26	MA	F	3.18	8	56	ST	F	1.73	11	86	PH	F	1.75	16
27	MA	F	3.67	8	57	ST	M	1.36	10	87	PH	F	1.75	16
28	MA	F	1.49	8	58	ST	F	1.97	11	88	PH	F	1.81	16
29	MA	F	3.41	8	59	ST	F	1.47	11	89	PH	M	2.13	10
30	MA	F	1.86	7	60	ST	M	1.63	11	90	PH	M	2.31	12

قم بإدخال البيانات في ملف جديد في برنامج SPSS وحفظه باسم "معدلات الطلبة"، (مستخدما القيم 1، 2، و3 لترميز الأقسام MA، ST، وPH على الترتيب، والقيم 1 و2 لترميز الذكور والإناث على الترتيب)، بعد ذلك قم من شريط الأدوات العلوي باختيار Analyze>Descriptive Statistics>Explore، وستظهر لك نافذة جديدة كما يوضح الشكل (45.3).

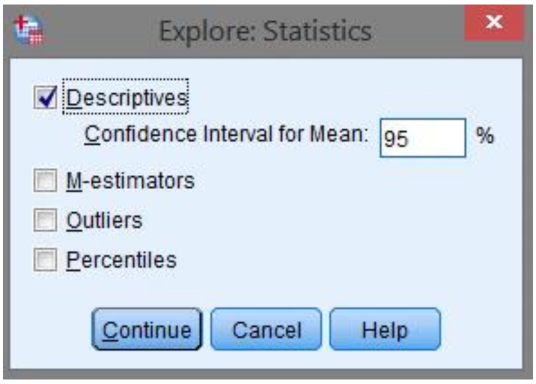


شكل 45.3: نافذة استكشاف البيانات (Explore) للملف "معدلات الطلبة".

في هذه النافذة، يمكنك اختيار المتغيرات التي ترغب في استكشافها، (مثل معدل_الطالب والفصل_الدراسي)، أي يمكنك حساب المقاييس الإحصائية والرسومات البيانية الأساسية لها عن طريق اختيارها ونقلها إلى مربع القائمة الأساسية (Dependent List)، بعد ذلك يمكنك الضغط على خيار الإحصاءات (Statistics) لاختيار النتائج المطلوبة ثم خيار الرسومات (Plots) لاختيار الرسومات البيانية المطلوبة، مع ملاحظة ترك خيار الاثنين (Both) في أسفل النافذة كما هو.

في هذا المثال، سنترك للقارئ استكشاف بيانات الطلبة باستخدام الخطوات السابقة، وسنقوم باستخدام أمر الاستكشاف (Explore) عن طريق عرض النتائج مقسمة أو مصنفة بحسب تخصص الطالب (القسم_العلمي) وذلك لغرض مقارنة معدلات الطلبة القدامى بين الأقسام.

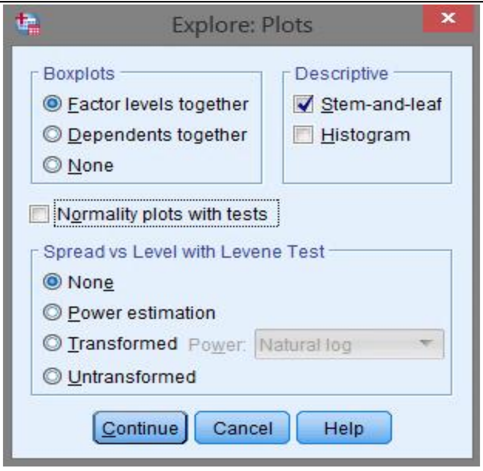
من جديد، قم باختيار المتغير "معدل_الطالب" في نافذة (Explore) ونقله إلى مربع القائمة (Dependent List)، بعد ذلك قم باختيار متغير التصنيف "القسم_العلمي" ونقله إلى مربع قائمة التصنيف (Factor List). الآن قم بالضغط على خيار الإحصاءات (Statistics) فتظهر نافذة فرعية كما يظهر في الشكل (46.3)؛



شكل 46.3: نافذة خيار الإحصاءات ضمن نافذة الاستكشاف (Explore).

ستلاحظ في تلك النافذة اختيار مسبق لأمر "وصفي" (Descriptive) والذي سنتركه كما هو، وتلاحظ أسفل منه وجود مربع خيار نسبة الثقة لفترة الثقة للوسط الحسابي للمجتمع (Confidence Interval for Mean) وهي النسبة الافتراضية 95% والتي سنتركها أيضا كما هي. قم بالنقر على المربع المقابل لخيار المئينات (Percentiles) والذي يوفر بعض المقاييس الهامة لتوزيع المتغير كما سنوضح تاليا، ولاحظ أنه يمكنك أيضا اختيار إظهار أبرز القيم المتطرفة للمتغير في جدول خاص إذا ما قمت بالنقر على الخيار الخاص بها (Outliers)، وهو ما سنقوم بالتعرض له في البند الأخير من هذا الفصل، وسنعتمد هنا على أسلوب آخر لاكتشاف القيم المتطرفة كما سنرى. الآن اضغط استمرار (Continue) في هذه النافذة، وستعود للنافذة السابقة، في الشكل (45.3)، قم بعدها بالضغط على خيار الرسم (Plot) وستظهر نافذة جديدة كما هو موضح في الشكل التالي، شكل (47.3)؛

في المربع الأيمن من النافذة سنقوم باختيار عرض المدرج التكراري (Histogram) إضافة لخيار عرض رسم الساق والورقة (Stem and leaf). وسنقوم أيضا باختيار عرض رسومات واختبارات التوزيع الطبيعي (Normality plots with Test) في الخانة التي تليها، ثم نضغط استمرار بعد ذلك فنرجع لنافذة الاستكشاف السابقة والتي سنضغط فيها على موافق ليتم تنفيذ كل ما تم اختياره في النوافذ الفرعية؛ نافذة الإحصاءات ونافذة الرسم.



شكل 47.3: نافذة خيار الرسم ضمن نافذة الاستكشاف.

في نافذة المخرجات، والتي نقترح تخزينها في ملف مخرجات خاص باسم "استكشاف معدلات الطلبة"، ستلاحظ ظهور عدد كبير من الجداول والرسومات البيانية والتي سنفرد البند القادم للتعليق عليها.

6.3 تفسير نتائج استكشاف البيانات (Interpreting the Results of Data Exploration)

تناولنا فيما سبق طرق متعددة لاستكشاف البيانات أو المتغيرات بصورة أحادية، (أي حساب المقاييس الإحصائية واستخدام التمثيل البياني لكل متغير على حده)، وهنا سنستعرض كيفية تفسير أهم النتائج الخاصة باستكشاف المتغيرات الأحادية، وأيضا الاستكشاف المتضمن للتصنيف بحسب متغير وصفي.

في الفقرات السابقة لهذا البند، تناولنا المثال الخاص بمعدلات الطلبة القدامى، (ملف "معدلات الطلبة")، وتم استخدام أمر استكشاف (Explore) ضمن قائمة (Analyze>Descriptive Statistics) وتعريف المتغير "معدل_الطالب" كمتغير كمي نرغب باستكشاف طبيعته وتوزيعه، ومتغير "القسم_العلمي" كمتغير وصفي يُستخدم لتصنيف معدلات الطلبة بناء على التخصص؛ رياضيات (MA)، إحصاء (ST)، وفيزياء (PH)، وقد اقترحنا تخزين المخرجات باسم "استكشاف معدلات الطلبة".

في نافذة المخرجات هذه، سيكون عرض النتائج بالصورة التالية؛

1. البداية تكون عادة بجداول النتائج، والجدول الأول، الذي هو بعنوان ملخص معالجة المشاهدات (Case Processing Summary)، يتم فيه عرض عدد ونسب المشاهدات في كل قسم، ونلاحظ أنها 30 مشاهدة لكل قسم بمجموع 90 مشاهدة كلية ولا توجد قيم مفقودة.

2. يضم الجدول الثاني، وهو بعنوان الإحصاءات الوصفية (Descriptives)، المقاييس الإحصائية لمتغير "معدلات_الطلبة" مقسمة بحسب التخصص أي بحسب القسم العلمي للطالب بحيث يضم الجزء الأول من الجدول

نتائج طلبة قسم الرياضيات (MA)، والجزء الثاني نتائج طلبة قسم الفيزياء (PH)، والجزء الثالث نتائج طلبة قسم الإحصاء (ST)، وهذا الترتيب في عرض النتائج هو بحسب الأبجدية الإنجليزية للتقسيمات MA ثم PH ثم ST.

من هذه المقاييس يمكننا استكشاف التالي:

- من قيم الوسط الحسابي (Mean) للطلبة في الأقسام يمكن القول بأن مستوى طلبة قسم الرياضيات هو أعلى من القسمين الآخرين. ولاحظ أن القيمة على يمين الوسط الحسابي هي الخطأ المعياري له ($SE(\bar{X}_{MA})$).
- بناء على ارتفاع قيمة الوسط الحسابي لمعدل طلبة قسم الرياضيات، فإن تقدير الوسط الحسابي لمجتمع معدلات طلبة قسم الرياضيات يتراوح في فترة ثقة أكبر من القسمين الآخرين، $1.75 < \mu_{MA} < 2.25$، أي أننا واثقون بنسبة 95% بأن معدلات طلبة قسم الرياضيات ستتراوح في العموم بين حدي الفترة السابقين. أما بالنسبة لمعدلات طلبة قسمي الفيزياء والإحصاء فإن تقدير متوسطاتها في العموم سيتراوح في مجال أضيق، وهو $1.61 < \mu_{PH} < 1.76$ و $1.55 < \mu_{ST} < 1.74$ تقريبا على التوالي.
- بمقارنة قيم الوسيط (Median) بقيم الوسط الحسابي لمعدلات الطلبة في كل قسم، نلاحظ أن معدلات الطلبة في قسم الرياضيات تحتوي على قيم متطرفة عليا بشكل ملاحظ، $\bar{X}_{MA} = 2.00 > \tilde{X}_{MA} = 1.79$ أما في قسمي الفيزياء والإحصاء فقيم الوسيط والوسط الحسابي متقاربة من بعضها البعض.
- بالنظر لقيم الانحراف المعياري (Std. Deviation) وقيم المدى (Range) لمعدلات الطلبة في الأقسام، نستطيع القول بأن معدلات طلبة قسم الرياضيات الدراسية هي الأكثر اختلافا أو تشتتا، ($S.D_{MA} = 0.67$، $Range_{MA} = 2.41$)، مقارنة بطلبة قسمي الفيزياء والإحصاء. وبالنظر إلى المدى الربيعي (Interquartile Range) فيمكن ملاحظة أن معدلات طلبة قسم الإحصاء هي الأكثر تجانسا. ويؤيد هذه الملاحظة القيم الصغرى والكبرى (Minimum and Maximum) لمعدلات الطلبة في كل قسم.
- قيم الالتواء (Skewness) تشير إلى وجود التواء حاد إلى اليمين في توزيع معدلات طلبة قسم الرياضيات يليه قسم الفيزياء لكن بالتواء أقل حده، أما بالنسبة لتوزيع معدلات قسم الإحصاء فيوجد التواء طفيف جدا إلى اليسار.

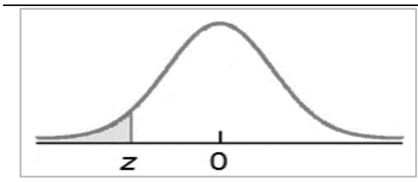
3. الجدول الثالث بعنوان المئينات (Percentiles) يضم مقاييس المئينات والربيعات (Quartiles) والتي تعتبر من المقاييس الهامة في اكتشاف توزيع المتغيرات. فمثلا؛ القيم للمئين 25 (الربيع الأول أيضا) للأقسام العلمية الثلاثة تُظهر بأن 25% من معدلات الطلبة هي أقل من المعدل 1.57 تقريبا (1.55 للرياضيات، 1.57 للفيزياء، و1.50 للإحصاء)، وهذا يعني أن ربع الطلبة مستواهم مقبول فأقل في الأقسام الثلاثة.

أما بالنسبة لقيم المئين 75 (وهي أيضا قيمة الربيع الثالث)، فإن 75% من طلبة قسمي الفيزياء والإحصاء معدلاتهم الدراسية هي مقبول فأقل، (1.75 و1.87 على التوالي)، وأما بالنسبة لقسم الرياضيات فإن 25% من الطلبة معدلاتهم هي أكبر من 2.13 أي بتقدير جيد، وهذا في المجمل يعني بأن مستوى طلبة قسم الرياضيات هو أفضل، ويمكن التعمق في هذا الاستنتاج من خلال مراقبة قيم المئينات الأخرى في الجدول.

4. الجدول التالي في ترتيب النتائج هو الجدول الخاص باختبار التوزيع الطبيعي، وهو بعنوان (Test of Normality). وهو يضم اختبارين شهيرين هما اختبار **شابيرو–ويلك** (Shapiro-Wilk Test) واختبار **كولموجوروف–سميرنوف** (Kolmogorov-Smirnov Test)، وكلاهما ويقوم باختبار الفرضية الصفرية التي تقضي بتبعية توزيع المتغير أو العينة بتوزيع طبيعي؛ **(العينة مسحوبة من مجتمع طبيعي :H_0)**.

وفي جدول النتائج نشاهد قيمة إحصاءة الاختبار (Statistic)، ودرجة الحرية (df)، والقيمة الاحتمالية (.Sig). وحيث أن قاعدة قبول أو رفض الفرضية الصفرية هي **(إذا كانت القيمة الاحتمالية (P-value) هي أقل من القيمة 0.05 نرفض الفرضية الصفرية)**[1]، فإنه يمكن القول من اختباري شابيرو–ويلك واختبار كولموجوروف–سميرنوف أن معدلات طلبة قسم الإحصاء تتبع التوزيع الطبيعي، ومعدلات طلبة قسم الفيزياء تتأرجح بين الطبيعية (اختبار كولموجوروف–سميرنوف) وغير الطبيعية (اختبار شابيرو–ويلك)، أما معدلات طلبة قسم الرياضيات فإنها لا تتوزع بتوزيع طبيعي بإجماع الاختبارين.

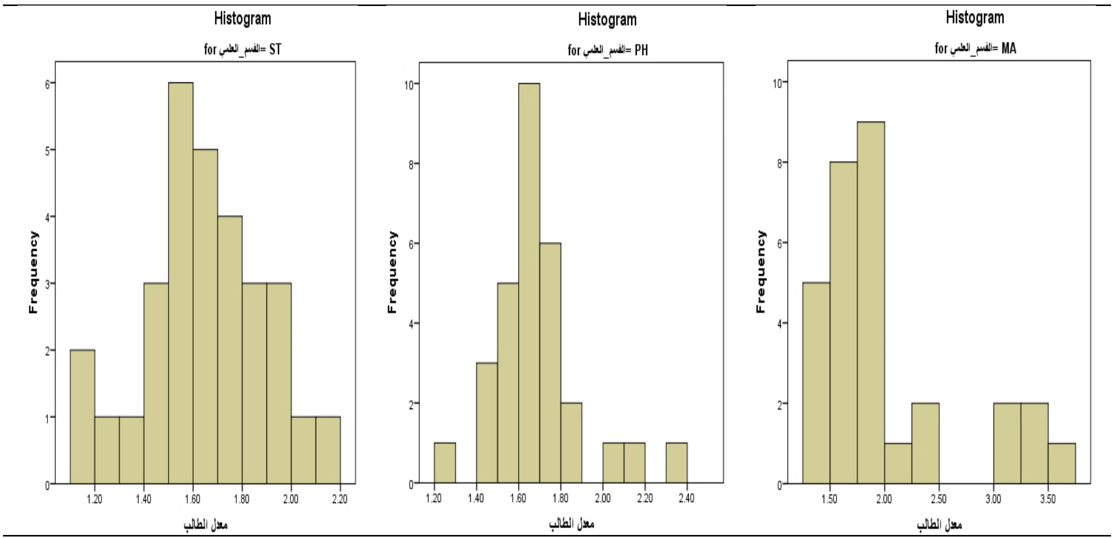
5. يلي الجداول في ملف المخرجات الرسومات البيانية، وأولها المدرجات التكرارية لمتغير الدراسة بحسب القسم العلمي، وبمقارنة المدرجات التكرارية الثلاثة نلاحظ أن المدرج الخاص بمعدلات قسم الإحصاء هو الأقرب لتشكيل الشكل الناقوسي المتماثل الشهير لمنحنى التوزيع الطبيعي، (شكل (48.3))، مع وجود التواء طفيف إلى اليسار.



شكل 48.3: منحنى التوزيع الطبيعي المعياري.

ومدرج معدلات قسم الفيزياء يوجد به التواء ظاهر إلى اليمين مع اقترابه إلى حد ما من التوزيع الطبيعي، أما المدرج التكراري الخاص بمعدلات قسم الرياضيات فهو أبعد ما يكون عن التوزيع الطبيعي مع وجود التواء كبير إلى اليمين في توزيع المشاهدات. وهذه النتائج متوافقة تماما مع ما سبق. ويمكن متابعة هذه الملاحظات من خلال الشكل (49.3).

[1] حيث أن 0.05 هي احتمال الوقوع في خطأ من النوع الأول، (رفض الفرضية الصفرية وهي صحيحة).



شكل 49.3: المدرجات التكرارية للمعدلات في أقسام الرياضيات، الفيزياء، والإحصاء، على الترتيب من اليمين إلى اليسار في ملف المخرجات "استكشاف معدلات الطلبة".

6. الرسومات التالية في الترتيب هي **رسومات الساق والورقة** (Stem-and-Leaf Plots) للأقسام العلمية. وهذه الرسومات هي وسيلة بصرية جيدة في التعرف على توزيع المتغيرات، وبالذات التوزيع الطبيعي في المتغيرات التي تضم مشاهدات قليلة نسبيا. وفكرة رسم الساق والورقة هي بسيطة بحد ذاتها، إذ يمثل العامود الأول إلى اليسار تكرارات قيم المشاهدات للمتغير، والعامود الثاني (الأوسط) الذي يمثل "الساق" والثالث الذي يمثل "الأوراق" هما القيم الفعلية للمشاهدات موزعة تصاعديا، (من أعلى الرسم إلى الأسفل)، بحيث يضم كل صف القيم القريبة من بعضها البعض.

ملاحظة:

في مثالنا الحالي، نجد أن الساق قد يمثل خانة القيمة الصحيحة للمعدل بينما تمثل الأوراق القيمة الأولى بعد الفاصلة العشرية، (كما هو الحال في رسمي قسم الرياضيات والإحصاء)، أو قد يمثل الساق خانة القيمة الصحيحة للمعدل مع القيمة الأولى بعد الفاصلة العشرية، (كما نلاحظ في رسم قسم الفيزياء). فمثلا؛ في الرسم الأول الخاص بقسم الرياضيات نجد في الصف الأول تحت عامودي الساق والورقة الشكل:

22 . 1 والقيمة 2 تحت عامود التكرار، وهذا يتم قراءته كالتالي:

توجد مشاهدتان (معدلان) كلاهما يبدأ بالقيمة 1.2، (وهاتين القيمتين هما فعليا 1.26 و 1.26 ضمن المشاهدات)، ولهذا تكون قيمة التكرار 2، وهكذا الأمر بالنسبة لباقي الصفوف في الرسم.

ولملاحظة ما إذا كان رسم الساق والورقة يعكس توزيعا طبيعيا للمتغير، يمكنك اعتبار القيم في صفوف الساق والورقة بمثابة الأعمدة في المدرج التكراري، فبالنظر إلى رسم الساق والورقة لقسم الإحصاء يمكنك ملاحظة امتداد القيم في الصفوف الوسطى وانحسارها في الصفوف العليا والسفلى، وهذا يشبه التوزيع الذي يمثله منحنى التوزيع الطبيعي ولكن بصورة مقلوبة. والميزة الأساسية في رسم الساق والورقة عن باقي الرسومات أنها توضح قيم

المشاهدات وتراكمها (تكرارها) بصورة فعلية واضحة. وهكذا يمكننا من خلال الرسومات الثلاثة للساق والورقة، (شكل (50.3))، ملاحظة اقتراب توزيع معدلات طلبة قسم الإحصاء من التوزيع الطبيعي بصورة أكبر من توزيع المعدلات في قسم الفيزياء، أما قسم الرياضيات فمعدلات طلبته تبتعد عن التوزيع الطبيعي كل البعد، وهذه النتائج مؤيدة لما سبقها من نتائج اختبارات الفروض والمدرجات التكرارية.

```
معدل الطالب Stem-and-Leaf Plot for
القسم_العلمي= MA

 Frequency    Stem &  Leaf

     2.00        1 .  22
     7.00        1 .  4445555
     6.00        1 .  666777
     7.00        1 .  8888899
     1.00        2 .  0
     1.00        2 .  2
     1.00        2 .  4
     5.00 Extremes    (>=3.2)

 Stem width:      1.00
 Each leaf:        1 case(s)
```

```
معدل الطالب Stem-and-Leaf Plot for
القسم_العلمي= PH

 Frequency    Stem &  Leaf

     1.00 Extremes    (=<1.26)
     1.00       14 .  1
     2.00       14 .  56
     2.00       15 .  33
     3.00       15 .  778
     4.00       16 .  0123
     6.00       16 .  667788
     1.00       17 .  3
     5.00       17 .  55589
     2.00       18 .  14
     3.00 Extremes    (>=2.09)

 Stem width:      .10
 Each leaf:        1 case(s)
```

```
معدل الطالب Stem-and-Leaf Plot for
القسم_العلمي= ST

 Frequency    Stem &  Leaf

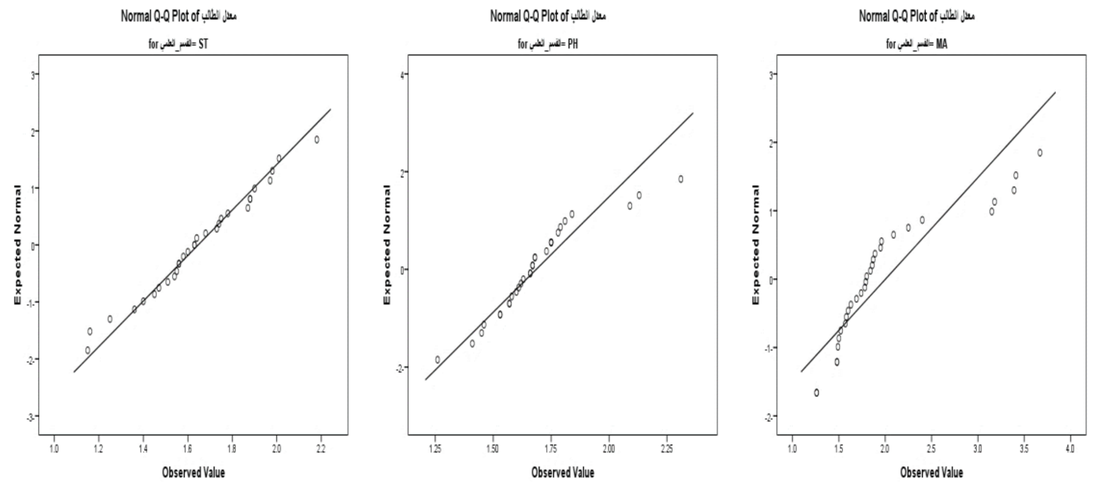
     2.00        1 .  11
     2.00        1 .  23
     9.00        1 .  444555555
     9.00        1 .  666667777
     6.00        1 .  888999
     2.00        2 .  01

 Stem width:      1.00
 Each leaf:        1 case(s)
```

شكل 50.3: شكل الساق والورقة للمعدلات في أقسام الرياضيات، الفيزياء، والإحصاء، على الترتيب من اليمين إلى اليسار في ملف المخرجات "استكشاف معدلات الطلبة".

إضافة إلى ما سبق، فإن رسم الساق والورقة يشكل أداة فعالة في التعرف على القيم المتطرفة في المتغير، فنلاحظ في قسم الرياضيات وجود خمسة معدلات متطرفة عليا عند النظر للصف الأخير في الرسم حيث نرى بأن التكرار يساوي 5، وظهور عبارة (Extremes (>=3.2)) والتي تدل على وجود قيم متطرفة وهي المعدلات التي هي أكبر من أو تساوي 3.2. وكذلك الأمر بالنسبة لقسم الفيزياء، حيث نرى وجود قيمة متطرفة صغرى (وهي 1.26)، وثلاثة قيم متطرفة عليا هي أكبر من أو تساوي المعدل 2.09. أما بالنسبة لمعدلات طلبة قسم الإحصاء فلا يُظهر الرسم أية قيم متطرفة.

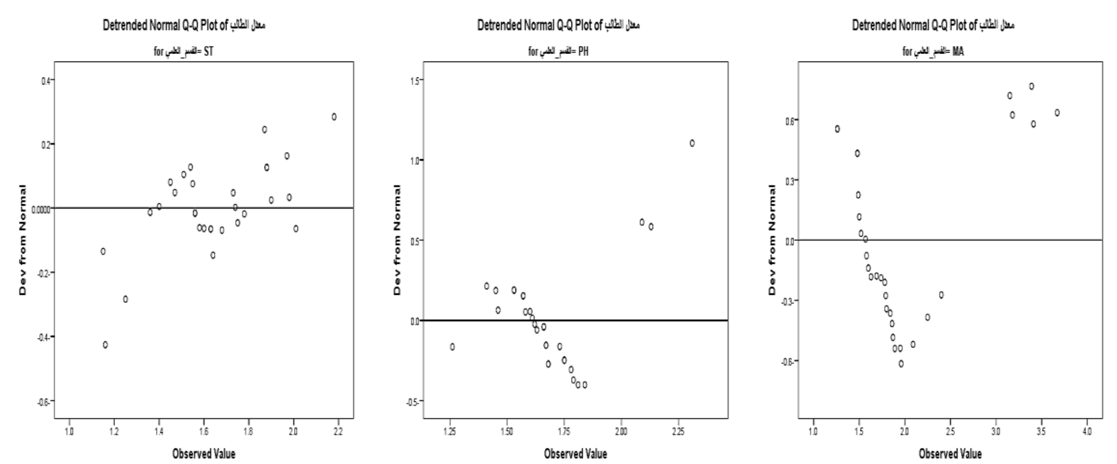
7. ننتقل الآن لرسم Q-Q، (شكل (51.3))، والذي تعرضنا له سابقا، ونلاحظ أن رسم Q-Q الخاص بمعدلات طلبة قسم الإحصاء هو الأقرب للتوزيع الطبيعي، أما النقاط على الشكل الخاص بقسم الفيزياء فتبتعد عن الخط المستقيم في القيم العليا، أما القيم في رسم معدلات قسم الرياضيات فتبتعد في المجمل عن الخط المستقيم، مما يدل على ابتعاد توزيع المشاهدات في هذا القسم عن التوزيع الطبيعي. وهذه النتائج تدور في نفس فلك التعليقات السابقة للأقسام العلمية الثلاثة.



شكل 51.3: رسم Q-Q للمعدلات في أقسام الرياضيات، الفيزياء، والإحصاء، في ملف المخرجات "استكشاف معدلات الطلبة".

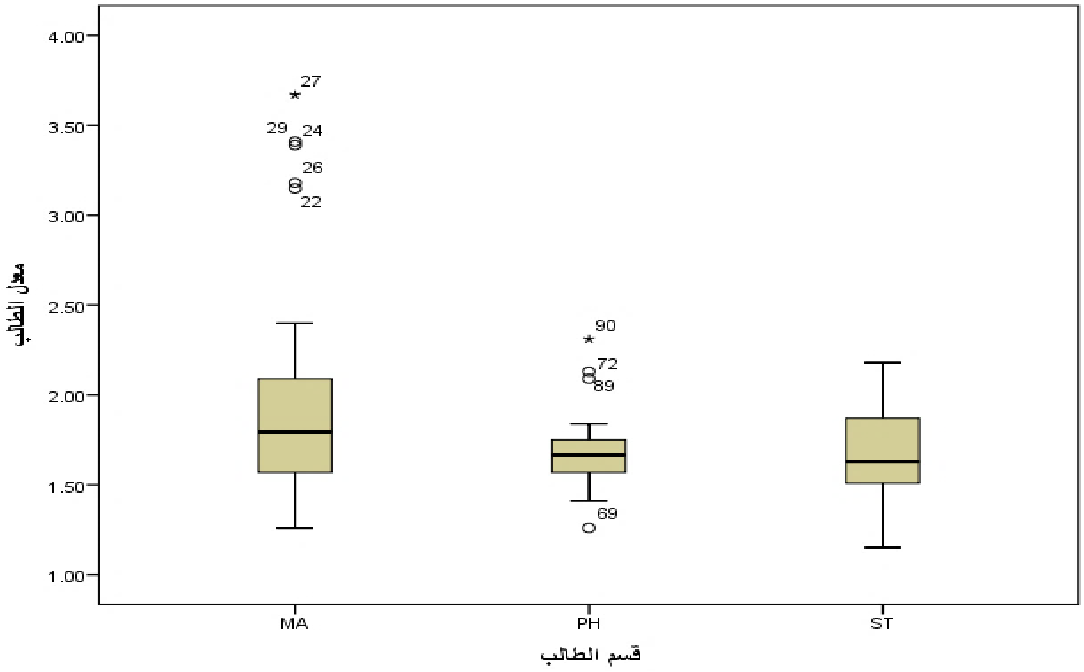
8. عندما يتم طلب رسم Q-Q في برنامج SPSS، فإنه يظهر عادة مرافقا لرسم آخر هو "نمط انتشار نقاط رسم Q-Q" (Detrended Normal Q-Q Plot). وهذا الرسم يمثل انتشار نقاط المشاهدات الفعلية مع قيمها المعيارية (Z Scores). فالمتغير الذي يتبع توزيعه توزيعا طبيعيا تكون النقاط فيه متجمعة أكثر ما يمكن حول الخط المستقيم الأفقي (عند القيمة صفر على المحور الصادي)، مما يعني اقتراب معظم قيم هذا المتغير من المركز، وبالتالي الاقتراب من التوزيع الطبيعي.

وفي مثالنا، نلاحظ، من شكل (52.3)، تجمع النقاط حول الخط الأفقي في الرسم الخاص بقسم الإحصاء، مما يدل على اقتراب توزيع معدلات الطلبة في هذا القسم من التوزيع الطبيعي. وفي الرسم الخاص بقسم الفيزياء، نلاحظ تجمع الكثير من النقاط حول الخط المستقيم وكذلك ابتعاد القيم العليا عنه مما يضع معدلات هذا القسم موضع الشك في مدى اقترابها من التوزيع الطبيعي. أما معدلات طلبة قسم الرياضيات فهي بعيدة جدا عن التوزيع الطبيعي وهذا ما يعكسه ابتعاد معظم النقاط عن الخط المستقيم على الرسم.



شكل 52.3: رسم نمط انتشار نقاط Q-Q للمعدلات في أقسام الرياضيات، الفيزياء، والإحصاء، على الترتيب من اليمين إلى اليسار في ملف المخرجات "استكشاف معدلات الطلبة".

9. الرسم الأخير في مثلنا، وهو شكل الصندوق (Boxplot)، تم عرضه لكل قسم علمي في شكل واحد، (شكل (53.3))، لتسهيل المقارنة بين توزيع المعدلات للأقسام.



شكل 53.3: شكل الصندوق للمعدلات في أقسام الرياضيات، الفيزياء، والإحصاء، على الترتيب من اليسار إلى اليمين، في ملف المخرجات "استكشاف معدلات الطلبة".

ويمكن الحصول على عدة استنتاجات من شكل الصندوق، من أهمها؛

- بالنظر إلى "طول" الصندوق المظلل لكل قسم، والذي يمثل توزيع أو انتشار المشاهدات[1]، يمكن ملاحظة أن انتشار البيانات (معدلات الطلبة) في قسم الرياضيات هو الأكبر، وهذا يدل على الاختلاف الكبير في المستوى الدراسي للطلبة القدامى في هذا القسم، وعلى النقيض يمكن ملاحظة تجانس المستوى الدراسي للطلبة القدامى في قسم الفيزياء إلى حد كبير، (نظرا "لانكماش" حجم الصندوق الخاص به). أما بالنسبة لقسم الإحصاء، فهنالك تجانس إلى حد ما في المستوى الدراسي. وهذه الاستنتاجات قد تم الوصول لها سابقا بطريقة رقمية من خلال حساب مقاييس التشتت لكل قسم علمي، (مثل الانحراف المعياري والمدى وغيرها).
- الخط الأفقي داخل كل صندوق مظلل، والذي يمثل قيمة الوسيط للمشاهدات لكل قسم، إضافة لارتفاع الصندوق المظلل ككل، يُظهر أن معدلات الطلبة في قسم الرياضيات هي الأعلى مقارنة بباقي الأقسام.

[1] إضافة لطول الأطراف، وهي الخطوط الخارجة من طرفي الصندوق المظلل لأعلى وأسفل، والتي تمثل القيم العليا والصغرى للمشاهدات ضمن سياق القيم المعتدلة، أي باستثناء القيم المتطرفة.

- النقاط المرقمة أعلى و/أو أسفل الصندوق تمثل القيم المتطرفة في المشاهدات، وهذه الأرقام تمثل ترتيب هذه القيم في المتغير. ويمكن بوضوح رؤية أن قسم الرياضيات يحتوي على خمسة قيم متطرفة عليا، (أكبرها هي القيم السابعة والعشرون في المتغير، وهو المعدل 3.67 في ذلك القسم). وأما معدلات الطلبة في قسم الفيزياء، فتضم ثلاثة قيم متطرفة عليا، (أكبرها القيمة التسعون، وهو المعدل 2.31 في ذلك القسم)، وقيمة متطرفة صغرى، (القيمة التاسعة والستين، وهو المعدل 1.26 في ذلك القسم). وأما قسم الإحصاء فلا تظهر فيه معدلات متطرفة للطلبة القدامى.

وبهذا نكون قد استعرضنا أهم الاستنتاجات التي تم الحصول عليها من خلال استخدام بعض أدوات الإحصاء الاستكشافي الأساسية، وقد تكون من أبرز ملاحظات القارئ هي تكرار الحصول على نفس النتائج من خلال استخدام مقاييس إحصائية أو رسومات بيانية مختلفة، وأن هذا قد يُعد إهدارا للوقت، إلا أننا نجيب هنا ببساطة إلى أن الحصول على نفس النتيجة أو الاستنتاج باستخدام أساليب مختلفة يُعد تأكيدا لهذا الاستنتاج ويُعطي الباحث المزيد من الثقة والثبات في نتائج تحليله الإحصائي.

وكتمرين إضافي للقارئ، نقترح إعادة تطبيق هذا البند على نفس المتغير "معدل_الطالب" في ملف البيانات "معدلات الطلبة" بطريقتين:

1. باستخدام المتغير "جنس_الطالب" كمتغير تقسيم بدلا عن المتغير "القسم_العلمي".
2. بدون إدراج متغير تقسيم، أي ترك مربع (Factor List) في نافذة استكشاف البيانات (Explore) في الشكل (45.3) فارغا.

7.3 استكشاف القيم المتطرفة (Exploring the Extreme Values)

من خلال استخدامنا للتمثيل البياني للمتغيرات مثل المدرج التكراري وشكل الصندوق في الأمثلة السابقة، لاحظنا ظهور بعض القيم المتطرفة في بعض المتغيرات، وفي هذا البند، سنقدم طريقة تحديد هذه القيم المتطرفة من خلال استخدام أحد خيارات أسلوب الاستكشاف في برنامج SPSS، والذي تمت الإشارة إليه سابقا؛

في ملف البيانات "معدلات الطلبة"، قم في شريط الأوامر العلوي باختيار Analyze>Descriptive Statistics>Explore فتظهر نافذة الاستكشاف، كما يظهر في الشكل السابق (شكل (45.3))، قم في تلك النافذة باختيار المتغيرين "معدل_الطالب" و"الفصل_الدراسي" ونقلهما إلى مربع القائمة الأساسية (Dependent List). بعد ذلك اضغط على خيار الإحصاءات (Statistics) في يمين النافذة فتظهر النافذة الفرعية الخاصة بالإحصاءات كما يظهر في الشكل السابق، (شكل (46.3)).

في تلك النافذة الفرعية سنقوم باختيار خيار القيم المتطرفة (Outliers) فقط ثم نضغط استمرار. وبعد العودة للنافذة الأساسية، سنختار عرض الإحصاءات فقط (Statistics) في الجزء الخاص بخيارات العرض (Display) في أسفل النافذة، ثم نضغط موافق.

في نافذة المخرجات، سنحصل على جدولين؛ الأول هو الجدول الخاص بأحجام العينات وعدد القيم المفقودة، وأما الجدول الثاني، (جدول (7.3) هنا)، فهو الجدول الخاص بالقيم المتطرفة. بالنسبة لمتغير معدل الطالب، فإن الجدول يُظهر قيم متطرفة عليا (Highest) وصغرى (Lowest) لهذا المتغير، فمثلا؛ نلاحظ وجود خمسة قيم متطرفة عليا (حيث أن العامود الثاني من اليمين (Case Number) يمثل ترتيب القيم المتطرفة في المتغير، والعامود الأول من اليمين (Value) يعرض القيمة المتطرفة)، وهي 3.15، 3.18، 3.39، 3.41، و3.67.

جدول 7.3: القيم المتطرفة للمتغيرين "معدل_الطالب" و"الفصل_الدراسي" في ملف البيانات "معدلات الطلبة".

Extreme Values

			Case Number	Value
معدل الطالب	Highest	1	27	3.67
		2	29	3.41
		3	24	3.39
		4	26	3.18
		5	22	3.15
	Lowest	1	44	1.15
		2	42	1.16
		3	45	1.25
		4	69	1.26
		5	8	1.26[a]
الفصل الدراسي للطالب	Highest	1	61	23
		2	4	21
		3	63	21
		4	64	21
		5	1	20[b]
	Lowest	1	23	2
		2	53	5
		3	30	7
		4	25	7
		5	24	7

a. Only a partial list of cases with the value 1.26 are shown in the table of lower extremes.
b. Only a partial list of cases with the value 20 are shown in the table of upper extremes.

أما القيم المتطرفة الصغرى لهذا المتغير فهي أكثر من خمسة قيم حيث أن وجود الحاشية (a) فوق القيمة 1.26 يعني أنه قد تم عرض جزء فقط من القيم المتطرفة الصغرى لهذا المتغير. وهذا يعني أن الطلبة الذين تحصلوا على معدلات 3.15 فأكثر أو معدلات تتراوح حول المعدل 1.26 قد تم تصنيف معدلاتهم بأنها متطرفة بالنسبة لبقية الطلبة في الكلية.

وكذلك الأمر بالنسبة للمتغير الثاني، وهو الفصل الدراسي للطالب، فإن ترتيب الفصول الدراسية حول القيمة 20، (والمشار إليه بالحاشية (b))، قد تم تصنيفهم بأنهم اجتازوا الحد الطبيعي لعدد الفصول الدراسية في هذه الكلية، أي أنهم قد تأخروا في التخرج. وأما الطلبة الذين ترتيب فصولهم الدراسية 2، 5، و7 فقد تم تصنيفها كقيم متطرفة صغرى مقارنة بباقي الطلبة.

من جديد، يمكننا باستخدام نفس نافذة الاستكشاف عرض القيم المتطرفة لمتغير أو أكثر باستخدام متغير تصنيف وصفي أو أكثر كما سنرى؛

لنفس ملف البيانات "معدلات الطلبة"، قم في شريط الأوامر العلوي باختيار Analyze>Descriptive Statistics>Explore وفي نافذة الاستكشاف، (شكل (45.3))، قم في تلك النافذة باختيار المتغير "معدل_الطالب" ونقله إلى مربع القائمة الأساسية (Dependent List)، واختيار المتغيرين "القسم_العلمي" و"جنس_الطالب" ونقلهما إلى مربع قائمة التصنيف (Factor List).

بعد ذلك اضغط على خيار الإحصاءات (Statistics) في يمين النافذة وفي النافذة الفرعية الخاصة بالإحصاءات، (شكل (46.3))، قم باختيار خيار القيم المتطرفة (Outliers) فقط ثم اضغط استمرار. وبعد العودة للنافذة الأساسية، اختر عرض الإحصاءات فقط في الجزء الخاص بخيارات العرض في أسفل النافذة، ثم اضغط موافق.

في نافذة المخرجات، وإلى جانب الجداول الخاصة بأحجام العينات وعدد القيم المفقودة، سنلقي الضوء أولا على الجدول الخاص بالقيم المفقودة لمتغير معدل الطالب مصنفا بناء على القسم العلمي الذي يتبعه الطالب، (جدول (8.3))؛

جدول 8.3: القيم المتطرفة للمتغير "معدل_الطالب" مصنفا بحسب المتغير "القسم_العلمي"، (ملف البيانات "معدلات الطلبة").

Extreme Values

قسم الطالب				Case Number	Value
معدل الطالب	MA	Highest	1	27	3.67
			2	29	3.41
			3	24	3.39
			4	26	3.18
			5	22	3.15
		Lowest	1	8	1.26
			2	6	1.26
			3	11	1.48
			4	3	1.48
			5	28	1.49
	ST	Highest	1	34	2.18
			2	49	2.01
			3	33	1.98
			4	58	1.97
			5	37	1.90

(تابع) جدول 8.3: القيم المتطرفة للمتغير "معدل_الطالب" مصنفا بحسب المتغير "القسم_العلمي"، (ملف البيانات "معدلات الطلبة").

		Lowest	1	44	1.15
			2	42	1.16
			3	45	1.25
			4	57	1.36
			5	50	1.40
	PH	Highest	1	90	2.31
			2	89	2.13
			3	72	2.09
			4	81	1.84
			5	88	1.81
		Lowest	1	69	1.26
			2	64	1.41
			3	74	1.45
			4	68	1.46
			5	78	1.53[a]

a. Only a partial list of cases with the value 1.53 are shown in the table of lower extremes.

ومن هذا الجدول يمكننا ملاحظة التغيرات التي حدثت نتيجة استخدام متغير التصنيف "القسم_العلمي" للطالب، فالقيم المتطرفة العليا لمعدلات الطلبة أصبحت مختلفة في كل قسم بحيث أن؛

- معدلات الطلبة المتطرفة العليا تتراوح ما بين 3.15 و3.67، في قسم الرياضيات (MA)، وتتراوح ما بين 1.90 و2.18 في قسم الإحصاء (ST)، وتتراوح ما بين 1.81 و2.31 في قسم الفيزياء (PH).
- معدلات الطلبة المتطرفة الصغرى تتراوح ما بين 1.26 و1.49، في قسم الرياضيات (MA)، وتتراوح ما بين 1.15 و1.40 في قسم الإحصاء (ST)، وتتراوح حول 1.53 في قسم الفيزياء (PH).

وهذا يعني أن ما يُعد قيمة متطرفة في قسم معين قد لا يكون كذلك في قسم آخر، وهذا التصنيف قد يكون أكثر دقة في كثير من الأحيان في غياب التجانس في مستوى الطلبة بين الأقسام العلمية المختلفة. فمن النتيجة السابقة نستطيع القول بأن المعدلات المرتفعة "المتطرفة" للطلبة في قسم الرياضيات هي التي أدت لرفع الوسط الحسابي لمعدلات الطلبة في هذا القسم عن المعدلات في القسمين الآخرين.

وثانيا، بالنسبة للجدول الآخر في نافذة المخرجات، (جدول (9.3))، فيمكن ملاحظة أن؛

- معدلات الطلبة المتطرفة العليا تتراوح ما بين 2.09 و2.40 للطلبة الذكور، وتتراوح ما بين 3.15 و3.67 للطالبات.
- معدلات الطلبة المتطرفة الصغرى تتراوح ما بين 1.26 و1.51 للذكور وتتراوح ما بين 1.15 و1.26 للطالبات.

وهذا يسوف يعطي الانطباع بأن المستوى الدراسي للطالبات في العموم هو أعلى من الطلبة الذكور لأن القيم المتطرفة العليا ستؤدي لارتفاع الوسط الحسابي لمعدلات الطالبات.

جدول 9.3: القيم المتطرفة للمتغير "معدل_الطالب" مصنفا بحسب المتغير "جنس_الطالب"، (ملف البيانات "معدلات الطلبة").

Extreme Values

جنس الطالب				Case Number	Value
معدل الطالب	M	Highest	1	17	2.40
			2	90	2.31
			3	34	2.18
			4	89	2.13
			5	18	2.09
		Lowest	1	8	1.26
			2	57	1.36
			3	50	1.40
			4	47	1.45
			5	32	1.51
	F	Highest	1	27	3.67
			2	29	3.41
			3	24	3.39
			4	26	3.18
			5	22	3.15
		Lowest	1	44	1.15
			2	42	1.16
			3	45	1.25
			4	69	1.26
			5	6	1.26

وعادة ما نقوم "بربط" النتائج الخاصة بالقيم المتطرفة بالنتائج العامة إذا ما تبين لنا وجود تأثير هام لتلك القيم ضمن متغيرات الدراسة، حيث أن زيادة عدد وقيمة التطرف في مشاهدات المتغير تؤدي لتغير النمط الذي يُظهره هذا المتغير.

الفصل الرابع

تحليل البيانات الاستدلالي في SPSS
(Inferential Data Analysis in SPSS)

في الفصل السابق تناولنا أهم الطرق الإحصائية من مقاييس وتمثيل بياني لاستكشاف ووصف البيانات بغية الوصول لفهم واضح حول توزيع هذه البيانات وتحديد الأنماط التي تتخذها وتحويل ذلك إلى معلومات "تمهيدية" يمكن الاستفادة منها. وفي هذا الفصل، سنقوم باستعراض الأساليب الإحصائية التي تتضمن طرق تقدير المعالم واختبارات الفروض المتعلقة بها، ويشمل ذلك تقدير معالم النماذج الإحصائية المختلفة، وذلك بهدف الوصول للمعلومات الكامنة ضمن طيات البيانات بصورة أكثر عمقا مما تم الوصول له بالأساليب الاستكشافية. وتندرج كل تلك المواضيع تحت مفهوم علم الإحصاء الاستدلالي (أو الاستنتاجي)، ذلك العلم الذي يتكون من مجموعة من الطرق التي تستخدم في استنتاج التقديرات والوصول لدلالات واختبار الفرضيات الإحصائية المتعلقة بمجتمع ما أو أكثر.

1.4 الاستدلال حول الوسط الحسابي للمجتمع (Inference about the Population Mean)

يُقصد بالاستدلال حول الوسط الحسابي للمجتمع هنا تقدير معلمة المجتمع بنقطة أو بفترة، واختبار مساواتها لقيمة مفترَضة، ويندرج هذا الموضوع في كثير من كتب الاستدلال الاحصائي تحت عنوان تقدير أو اختبار العينة الواحدة والعينتين. وسنبدأ بتناول المجتمع الواحد.

1.1.4 الاستدلال حول الوسط لمجتمع واحد (Inference about the Mean of One Population)

توجد عدة صيغ لتقدير فترة الثقة للوسط الحسابي للمجتمع μ وإجراء الاختبارات المتعلقة به، وهذه الصيغ تختلف باختلاف حجم العينة ومعلومية تباين المجتمع، لذلك سيتم عرض الصيغة النظرية لفترة الثقة بمعلومية تباين المجتمع σ^2؛

يتم تعريف $100(1-\alpha)\%$ فترة ثقة للوسط الحسابي μ بالصيغة التالية:

$$\mu_X: \bar{X} \mp Z\alpha_{/2} \frac{\sigma}{\sqrt{n}}$$

حيث $\bar{X}$ هو الوسط الحسابي لعينة حجمها n مسحوبة من هذا المجتمع، و $Z\alpha_{/2}$ هي قيمة الإحصاءة Z التي تتبع التوزيع الطبيعي المعياري باحتمال قدره $^{\alpha}/_{2}$.

ويتم اختبار الفرضية الصفرية (أو فرضية العدم):

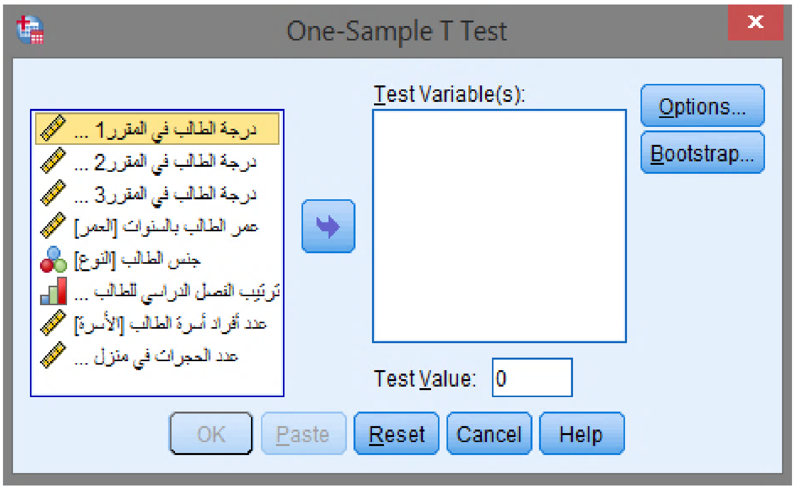
$$H_0: \mu = \mu_0$$

عند مستوى معنوية α باستخدام الإحصاءة Z_c بالصورة:

$$Z_c = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

ولعمل استدلال حول الوسط الحسابي لمجتمع في برنامج SPSS، نقوم أولا بفتح ملف البيانات، وليكن ملف "بيانات الطلبة" على سبيل المثال. في هذا الملف، لنفرض أننا نريد أولا تقدير فترة الثقة لمجتمع أعمار الطلبة في تلك الكلية، واختبار فرضية حول قيمة معينة لمتوسط المجتمع (معلمة المجتمع)، فنقوم بالآتي؛

في شريط الأدوات العلوي اختر الأمر Analyze>Compare Means>One-Sample T Test، فتظهر النافذة التالية، شكل (1.4)؛



شكل 1.4: نافذة تقدير واختبار الوسط الحسابي لمجتمع واحد (لملف بيانات "بيانات الطلبة").

لتقدير الوسط الحسابي لمتغير العمر، $\mu_{\text{العمر}}$، قم باختيار المتغير "عمر الطالب بالسنوات" ونقله إلى مربع متغير الاختبار (Test Variable(s)). الآن، ومن أمر خيارات (Options) يمكنك تحديد نسبة الثقة التي ترغب بتقدير فترة الثقة عندها، وتكون عادة النسبة الافتراضية 95%، وإذا رغبت بإبقائها كما هي فلا داعي للضغط على أمر خيارات. بعد ذلك قم بتحديد القيمة التي سيتم اختبار المعلمة عندها، ولتكن 22 سنة على سبيل المثال، أي أننا سنختبر الفرضية الصفرية $H_0: \mu = 22$. قم بكتابة 22 في مربع قيمة الاختبار (Test Value)، ثم اضغط موافق (OK).

ستظهر في نافذة المخرجات النتيجة التالية، (الجدولين (1.4) و(2.4))، حيث يضم الجدول الأول حجم العينة، الوسط الحسابي للعينة ($\bar{X}$)، الانحراف المعياري للعينة ($SD(X)$)، والخطأ المعياري للمقدر ($SE(\bar{X})$). وهذه المقاييس يمكن الحصول عليها باستخدام بطرق أخرى كما رأينا في الفصل الثالث، إلا أنها تأتي هنا مرافقة لأمر تقدير فترة الثقة واختبار الفروض.

جدول 1.4: جدول الإحصاءات لعينة واحدة لمتغير "العمر" في ملف "بيانات الطلبة".

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
عمر الطالب بالسنوات	35	22.14	1.648	.278

أما الجدول الثاني، (جدول (2.4))، فيضم قيم إحصاءة الاختبار [1] ودرجة الحرية لها، والقيمة الاحتمالية للاختبار (P-value) وهي للاختبار من طرفين، (حيث لا يتوفر هنا خيار إجراء الاختبار من طرف واحد)، وقيمة الفرق ($\bar{X} - \mu_0$)، وفترة الثقة للوسط الحسابي.

جدول 2.4: جدول اختبار الوسط الحسابي لعينة واحدة لمتغير "العمر" في ملف "بيانات الطلبة".

One-Sample Test

	Test Value = 22					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
عمر الطالب بالسنوات	.513	34	**.611**	.143	-.42-	.71

ومن هذه النتيجة، يمكن القول أننا نميل لقبول الفرضية الصفرية (حيث أن P-value > 0.05) بنسبة ثقة 95%، مما يعني قبول الادعاء القائل بأن متوسط أعمار الطلبة في المجتمع هو 22 سنة، وأن تقدير القيمة الحقيقية (المعلمة) لهذا المتوسط سيتراوح ما بين (21.576≅22) و(22.708≅23) سنة.

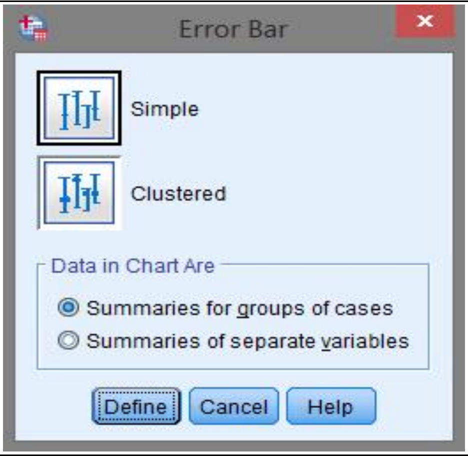
1.1.1.4 استخدام أعمدة الخطأ (Using the Error Bars)

لنفرض بأننا نرغب في المثال السابق بالمقارنة بين فترات الثقة لمتوسط مجتمع درجات الطلبة للمقررات (المتغيرات) الثلاثة "مقرر1"، "مقرر2"، و"مقرر3" في ملف البيانات "بيانات الطلبة". عندها يمكن استخدام الرسم البياني **لأعمدة الخطأ** (Error Bars) في تنفيذ ذلك. من شريط الأوامر العلوي نقوم باختيار الأمر Graphs>Legacy Dialogs>Error Bar فتظهر نافذة الرسم كما في شكل (2.4).

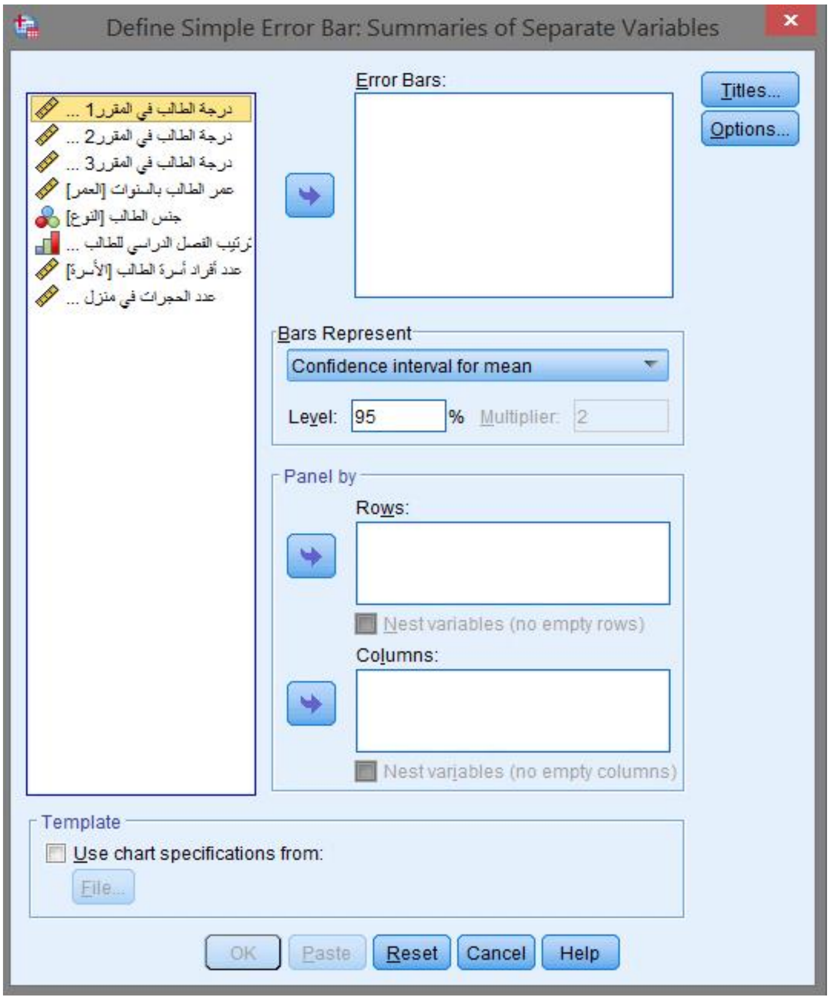
[1] ننوه هنا إلى أنه في معظم البرامج الإحصائية يتم إطلاق مسمى اختبار t على الاختبارات التي تتضمن استخدام كلا من التوزيع الطبيعي Z وتوزيع t حيث أن قيم الإحصاءة في التوزيع الثاني تقترب من التوزيع الطبيعي بزيادة حجم العينة.

في تلك النافذة، سنقوم بإبقاء اختيار الأعمدة البسيطة (Simple) كما هو ولكن سنختار التنفيذ لعدة متغيرات (Summaries of separate variables)، ثم نضغط تعريف (Define) فتظهر نافذة جديدة كما في الشكل (3.4).

في هذه النافذة، سنقوم باختيار المتغيرات التي نرغب بتنفيذ أعمدة الخطأ لها، (وهي المتغيرات الثلاثة الأولى في قائمة المتغيرات)، ونقلها إلى مربع أعمدة الخطأ (Error Bars) في يمين النافذة.



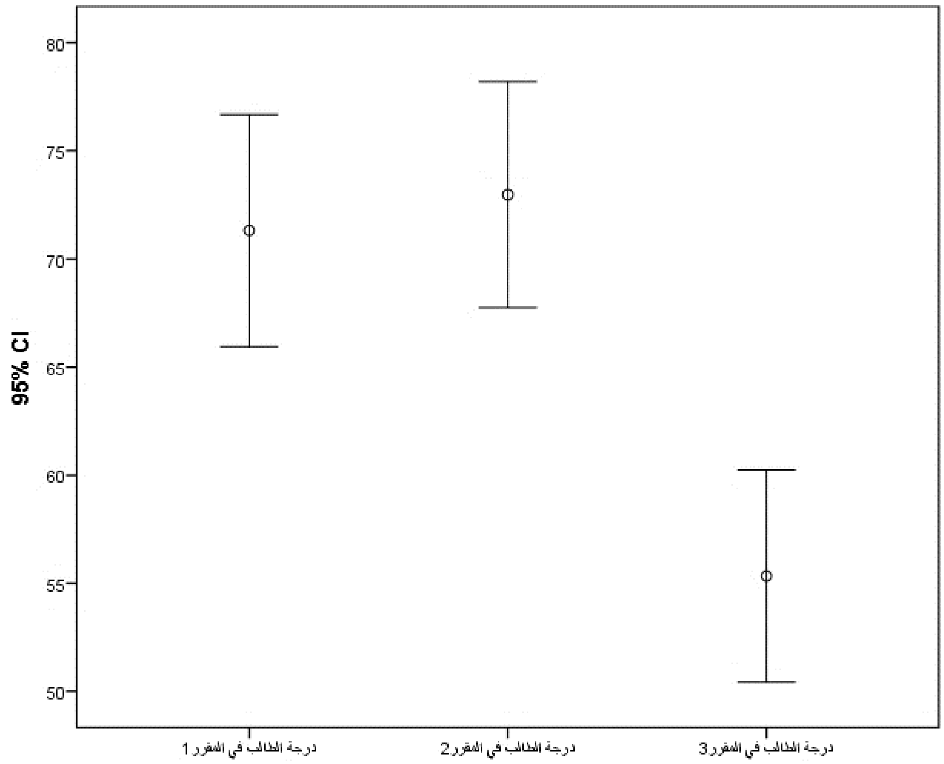
شكل 2.4: نافذة أعمدة الخطأ (Error Bars).



شكل 3.4: نافذة تعريف المتغيرات لرسم أعمدة الخطأ (لملف بيانات "بيانات الطلبة").

وتحت مربع أعمدة الخطأ، يوجد خيار خاص بما ستمثله هذه الأعمدة، (Bars Represent)، وإذا ما ضغطت على السهم الخاص به ستظهر ثلاثة خيارات هي؛ فترة الثقة للوسط، الخطأ المعياري للوسط، والانحراف المعياري. في مثالنا سنبقي على الخيار الأول الافتراضي، وكذلك سنبقي على اختيار %95 نسبة ثقة للفترة، ونضغط موافق في أسفل النافذة. سيظهر بعدها رسم أعمدة الخطأ للمتغيرات الثلاثة المختارة في نافذة المخرجات، كما نشاهد في الشكل (4.4). من الشكل يمكننا ملاحظة ارتفاع مستوى أداء الطلبة في المقررين 1 و2 مقارنة بأدائهم في المقرر3، (لاحظ أن النقطة المستديرة في منتصف كل عامود تمثل متوسط العينة).

وأيضا يمكننا القول إنه في العموم، وكتقدير لمتوسط المجتمع الذي يمثل درجات الطلبة في المقررات الدراسية، أن مستوى الطلبة في المقررين 1 و2 سيتراوح بين جيد وجيد جدا، (من 66 إلى 77 درجة تقريبا)، وأما في المقرر3 فسيتراوح ضمن درجات التقدير مقبول.

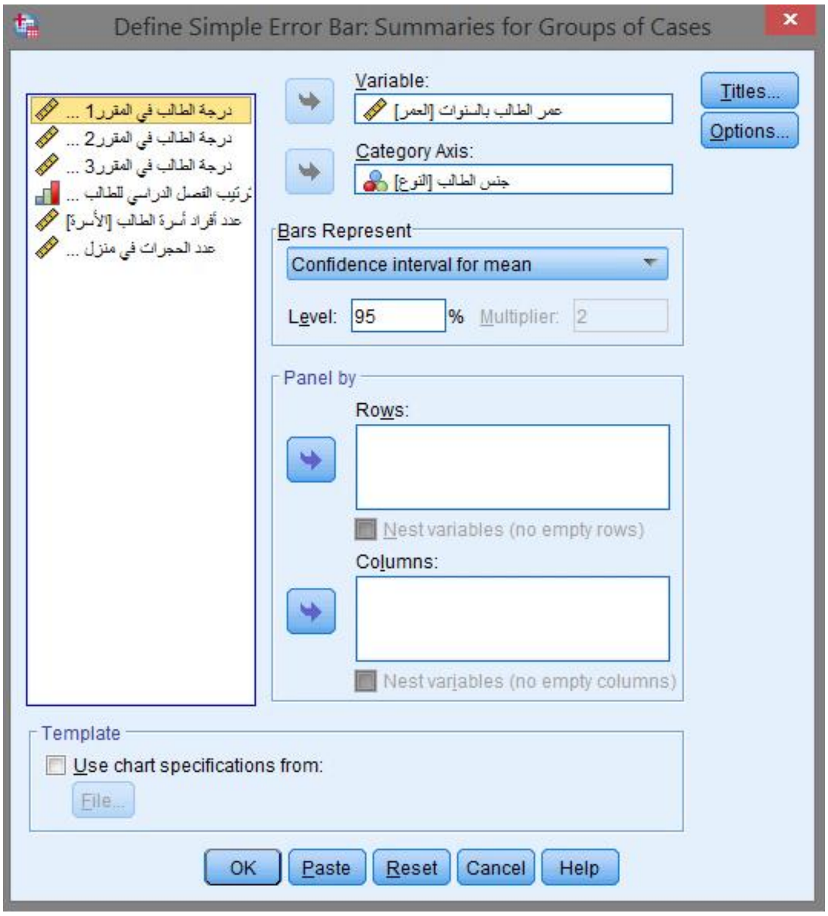


شكل 4.4: التمثيل البياني لأعمدة الخطأ للمتغيرات "مقرر1"، "مقرر2"، و"مقرر3"، (لملف بيانات "بيانات الطلبة").

ويمكن أيضا استخدام أعمدة الخطأ للمقارنة بين فترات الثقة لمتغير واحد بناء على متغير تقسيم وصفي، فمثلا؛ يمكننا تقدير ورسم فترتي ثقة لمتغير عمر الطالب، في نفس ملف البيانات "بيانات الطلبة" للذكور وللإناث ومقارنة الأداء الدراسي لكل منهما.

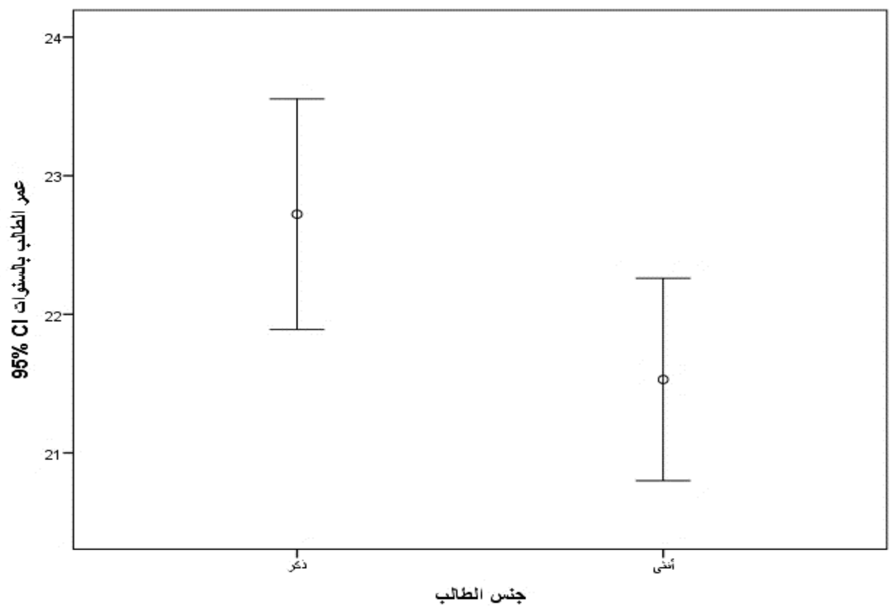
في شريط الأوامر العلوي قم باختيار الأمر Graphs>Legacy Dialogs>Error Bar، ثم اختر التنفيذ لمجموعات من المشاهدات (Summaries for groups of cases)، ثم اضغط تعريف (Define) فتظهر نافذة جديدة كما في الشكل (5.4).

في هذه النافذة، سنقوم باختيار تمثيل أعمدة الخطأ لمتغير عمر الطالب باستخدام متغير جنس الطالب كمتغير تقسيم. قم، كما هو موضح في الشكل (5.4)، باختيار المتغير "عمر الطالب بالسنوات" في مربع المتغير (Variable)، ثم قم باختيار المتغير "جنس الطالب" في مربع محور التصنيف (Category Axis)، مع إبقاء الخيار الافتراضي لفترة الثقة للوسط الحسابي، وخيار 95% نسبة ثقة للفترة كما هما.



شكل 5.4: نافذة اختيار تنفيذ أعمدة الخطأ لمجموعات من المشاهدات ضمن متغير باستخدام تقسيم لمتغير وصفي، (لملف بيانات "بيانات الطلبة").

بعد الضغط على موافق، سيظهر التمثيل البياني المطلوب لأعمدة الخطأ لعمر الطالب بناء على تقسيم المشاهدات بحسب جنس الطالب، (شكل (6.4)). ومن هذه الأعمدة يمكننا ملاحظة وجود ارتفاع ملحوظ في أعمار الطلبة الذكور مقارنة بالطالبات. كما يمكننا القول بأن تقدير الوسط الحسابي لمجتمع أعمار الطلاب الذكور سيكون في العموم أعلى من تقدير الوسط الحسابي لمجتمع أعمار الطالبات بنسبة ثقة 95%.



شكل 6.4: التمثيل البياني لأعمدة الخطأ لمتغير عمر الطالب بناء على تقسيم المشاهدات بحسب جنس الطالب، (لملف بيانات "بيانات الطلبة").

2.1.4 الاستدلال حول الوسط لمجتمعين (Inference about the Mean of Two Populations)

1.2.1.4 الاستدلال للعينات المستقلة (Inference about Independent Samples)

تُعرّف الصيغة العامة لإحصاءة اختبار الفرق بين متوسطين ($H_0: \mu_1 - \mu_2 = 0$) بالصورة التالية:

$$Z_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

وذلك في الحالات التي تتضمن العينات المستقلة الكبيرة أو عند توفر قيم التباين للمجتمعين، وتُستخدم إحصاءة t في الحالات الأخرى، (صغر حجم العينة أو عدم توفر معلومات حول تباينات المجتمعين)، والتي يجب أن يكون من المعلوم فيها تساوي أو عدم تساوي تباينات المجتمعين.

وتوجد طريقتين لإجراء استدلال حول الفرق بين الوسطين الحسابيين لمجتمعين في برنامج SPSS، الطريقة المختصرة والطريقة المطولة.

في **الطريقة المختصرة:** يتم إدخال قيم كل من حجم العينة، الوسط الحسابي للعينة، والانحراف المعياري للعينة وذلك لكل من العينتين مباشرة في نافذة التحليل كما سنوضح في المثال التالي؛

لنفرض أننا نريد إجراء استدلال إحصائي للفرق بين وسطي مجتمعات درجات الطلبة في المقررين 1 و2 في ملف البيانات "بيانات الطلبة"، أي نريد اختبار الفرضية الصفرية؛ $H_0: \mu_{1_{مقرر}} - \mu_{2_{مقرر}} = 0$، وتقدير 95% فترة ثقة للفرق بينهما. لتنفيذ ذلك نحتاج لحساب قيم الوسط الحسابي والانحراف المعياري للعينتين (أي للمتغيرين) أولا.

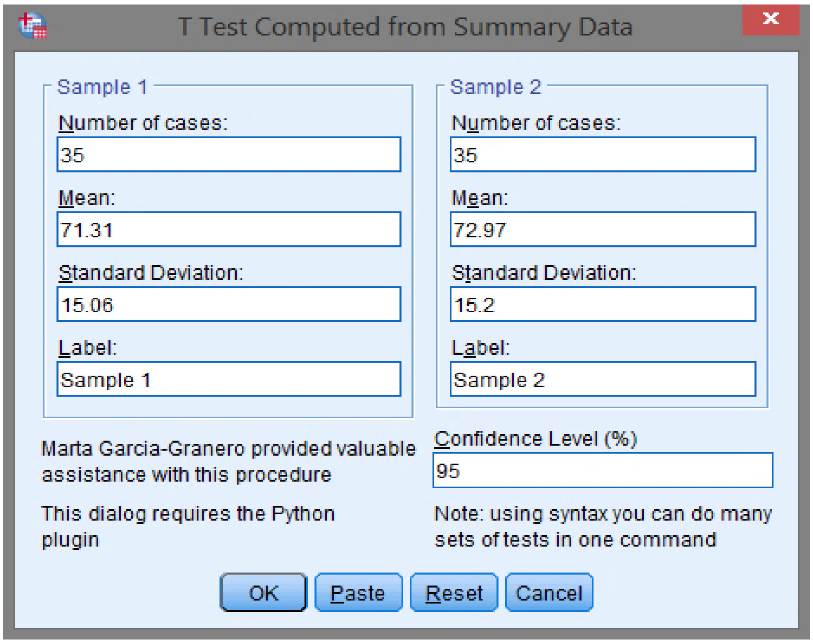
وبالتالي إما أن نكون قد حصلنا عل تلك القيم من التحليل الاستكشافي للمتغيرين مسبقا، أو نقوم بحساب تلك المقاييس مباشرة. في شريط الأوامر العلوي قم باختيار Analyze>Descriptive Statistics>Descriptives ثم اختيار المتغيرين "مقرر1" و"مقرر2" واختيار الوسط الحسابي والانحراف المعياري فقط من أمر الخيارات (Options) ثم الضغط على موافق فنحصل على القيم المطلوبة في نافذة المخرجات، (جدول (3.4)).

جدول 3.4: الوسط الحسابي والانحراف المعياري لمقرري 1 و2 في ملف البيانات "بيانات الطلبة".

Descriptive Statistics

	N	Mean	Std. Deviation
درجة الطالب في المقرر 1	35	71.31	15.605
درجة الطالب في المقرر 2	35	72.97	15.219
Valid N (listwise)	35		

من جديد، وفي شريط الأدوات العلوي، قم باختيار Analyze>Compare Means>Summary Independent Samples T-Test فتظهر نافذة كما في الشكل (7.4). قم بإخال القيم المطلوبة كما هو موضح في الشكل ثم اضغط موافق فتظهر النتائج ممثلة بثلاثة جداول في نافذة المخرجات، (جدول (4.4)).



شكل 7.4: نافذة إدخال القيم الخاصة بالاستدلال حول الفرق بين وسطي مجتمعين، (الطريقة المختصرة)، (لملف بيانات "بيانات الطلبة").

الجدول الفرعي الأول (أ) يضم قيم الوسط الحسابي والانحراف المعياري والخطأ المعياري للوسط للعينتين. ويضم الجدول الفرعي الثاني (ب) نتائج اختبار الفرق بين الوسطين في حالتين؛ الأولى بافتراض تساوي تبايني المجتمعين، (Equal variances assumed)، والثانية بافتراض عدم تساوي تبايني المجتمعين، (Equal variances not assumed).

جدول 4.4: نتائج الاستدلال حول الفرق بين الوسطين الحسابيين لمجتمعي المقررين 1 و 2 في ملف البيانات "بيانات الطلبة".

Summary Data (أ)

	N	Mean	Std. Deviation	Std. Error Mean
Sample 1	35.000	71.310	15.060	2.546
Sample 2	35.000	72.970	15.200	2.569

Independent Samples Test (ب)

	Mean Difference	Std. Error Difference	t	df	Sig. (2-tailed)
Equal variances assumed	-1.660	3.617	-.459	68.000	**.648**
Equal variances not assumed	-1.660	3.617	-.459	67.994	**.648**

Hartley test for equal variance: F = 1.019, Sig. = 0.4783

95.0% Confidence Intervals for Difference (ج)

	Lower Limit	Upper Limit
Asymptotic (equal variance)	-8.749	5.429
Asymptotic (unequal variance)	-8.749	5.429
Exact (equal variance)	-8.877	5.557
Exact (unequal variance)	-8.877	5.557

وفي أسفل الجدول الفرعي الثاني (ب)، نلاحظ وجود نتائج اختبار تساوي تبايني المجتمعين، (**اختبار هارتلي** (Hartley test for equal variance))، ومنه يتضح قبول الفرضية الصفرية $H_0: \sigma_1^2 = \sigma_2^2$، وبالتالي اعتماد نتائج اختبار الفرق بين وسطي المجتمعين الخاصة به في الصف الأول من الجدول. ومن هذه النتيجة نخلص إلى قبول الفرضية $H_0: \mu_{1\text{مقرر}} - \mu_{2\text{مقرر}} = 0$، (حيث أن P-value =0.648)، أي الاعتقاد بتساوي معدلات درجات الطلبة في مجتمعي المقررين 1 و2.

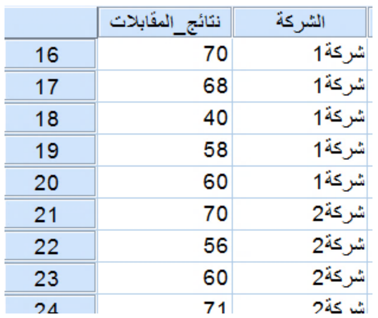
وأما الجدول الفرعي الثالث (ج)، فيضم فترة الثقة للفرق بين وسطي مجتمعي المقررين 1 و2، (بتقدير تقاربي (Asymptotic) وتقدير محدد (Exact))، ومنه يتضح أن $-8.87 < \mu_{1\text{مقرر}} - \mu_{2\text{مقرر}} < 5.55$ ، وهذا يعني أن درجات الطلبة في المقرر 1 ستكون في العموم أقل من درجات الطلبة في المقرر 2، وذلك بسبب كون الفرق بين التقديرين هو أقرب للقيمة السالبة (8.87−).

أما في **الطريقة المطولة:** فيجب إدخال العينيتين، (المطلوب تقدير الفرق بين متوسطات مجتمعيهما)، في متغير (عامود) واحد، وإدراج متغير وصفي يضم رمز التمييز بينهما، والمثال التالي يوضح هذه الطريقة. الجدول (5.4) يضم بيانات تمثل النتائج التي حصل عليها 20 شخصا أجروا مقابلات شخصية للعمل في قسم الموارد البشرية في شركتين، (حيث كان التقييم من 100 درجة)؛

وسنقوم بإدخال هذه البيانات في ملف جديد في SPSS باسم "المقابلات الشخصية"، عن طريق إدخال نتائج الشركة 1 متبوعة بنتائج الشركة 2 في عامود واحد، ثم إدخال رمز يميز بين نتائج الشركتين في العامود الثاني كما هو موضح في الشكل (8.4)، والذي يمثل مقطعا من البيانات، مع التنويه هنا إلى أنه تم إدخال القيمة 1 كرمز للشركة1، والقيمة كرمز 2 للشركة2 ثم تعريف ذلك في عامود القيم (Values).

جدول 5.4: نتائج المقابلات الشخصية في الشركتين.

	نتائج الشركة 1		نتائج الشركة 2
1	71	1	70
2	55	2	56
3	62	3	60
4	70	4	71
5	23	5	20
6	56	6	58
7	80	7	82
8	74	8	75
9	66	9	65
10	68	10	70
11	43	11	40
12	56	12	55
13	90	13	89
14	56	14	57
15	42	15	40
16	70	16	71
17	68	17	70
18	40	18	39
19	58	19	60
20	60	20	59

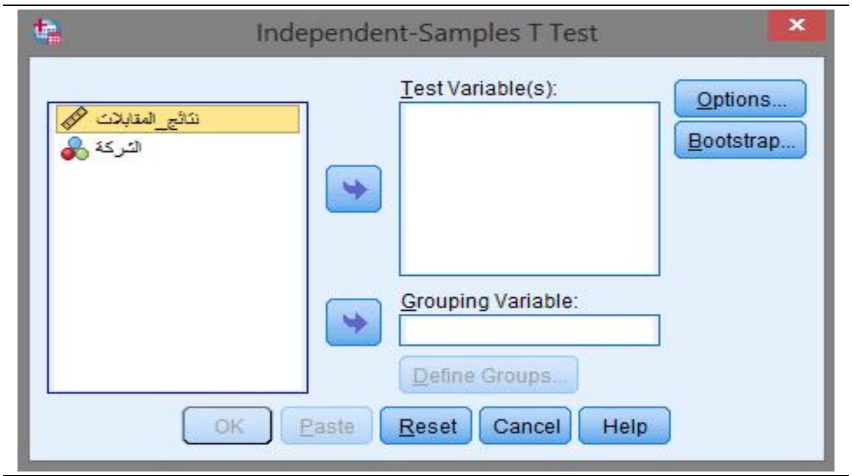


شكل 8.4: مقطع من ملف بيانات "المقابلات الشخصية".

ولإجراء استدلال إحصائي للفرق بين وسطي مجتمعات نتائج المقابلات الشخصية في الشركتين، أي اختبار الفرضية الصفرية؛

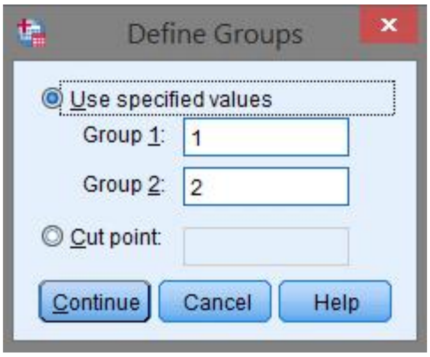
$H_0: \mu_{\text{شركة}1} - \mu_{\text{شركة}2} = 0$،

وتقدير 95% فترة ثقة للفرق بينهما، نقوم من شريط الأوامر العلوي باختيار Analyze>Compare Means>Independent Samples T-Test فتظهر النافذة الخاصة بتعريف المتغيرات كما في الشكل (9.4)؛ في تلك النافذة، قم باختيار المتغير "نتائج_المقابلات" ونقله لمربع متغير الاختبار (Test Variable)، ثم قم بنقل المتغير "الشركة" إلى مربع متغير التصنيف (Grouping Variable)، وعندها "سيضيئ" أسفل ذلك المربع الأمر الخاص بتعريف التصنيفات أو المجموعات (Define Groups)، قم بالضغط عليه لتعريف العينة الأولى والثانية.



شكل 9.4: نافذة إدخال القيم الخاصة بالاستدلال حول الفرق بين وسطي مجتمعين، (الطريقة المطولة)، (لملف بيانات "المقابلات الشخصية").

بعد الضغط على أمر تعريف المجموعات، ستظهر النافذة الفرعية كما في الشكل (10.4)، قم بكتابة القيمة 1 للمجموعة (العينة) الأولى والقيمة 2 للمجموعة (العينة) الثانية، ثم اضغط استمرار لتعود للنافذة الأصلية، (شكل (9.4))، اضغط بعدها على موافق فتظهر نتائج الاستدلال الإحصائي لتقدير الفرق بين وسطي مجتمعي نتائج الشركتين في نافذة المخرجات، (جدول (6.4)). في ذلك الجدول، ستلاحظ وجود قيم الوسط الحسابي للعينتين (الشركتين)، الانحراف المعياري للعينتين، والخطأ المعياري للوسط الحسابي للعينتين في الجدول الفرعي الأول (أ). وأما الجدول الفرعي الثاني (ب) فيحتوي على نتائج اختبار الفروض وتقدير فترة الثقة.



شكل 10.4: نافذة تعريف العينتين في ملف بيانات "المقابلات الشخصية".

جدول 6.4: نتائج الاستدلال حول الفرق بين وسطين لمجتمعي الشركتين 1 و2، (ملف بيانات "المقابلات الشخصية").

Group Statistics (أ)

	الشركة	N	Mean	Std. Deviation	Std. Error Mean
المقابلات_نتائج	شركة1	20	60.40	15.288	3.419
	شركة2	20	60.35	16.239	3.631

Independent Samples Test (ب)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
نتائج المقابلات	Equal variances assumed	.023	**.881**	.010	38	**.992**	.050	4.987	-10.046-	10.146
	Equal variances not assumed			.010	37.862	**.992**	.050	4.987	-10.047-	10.147

من الجدول الفرعي الثاني (ب) في الجدول (6.4)، يمكننا استنتاج التالي، (بقراءة أعمدة الجدول من اليسار لليمين):

- قبول الفرضية القائلة بتساوي تباينات مجتمعي الشركتين، (P-value =0.881)، وبالتالي اعتماد النتائج الخاصة بهذه الفرضية، $H_0: \sigma_1^2 = \sigma_2^2$ ، (Equal variances assumed).
- قبول الفرضية القائلة بتساوي الوسطين الحسابيين لمجتمعي الشركتين، (P-value =0.992)، مما يعني تساوي أداء الأشخاص في المقابلات الشخصية في كلا الشركتين.
- التقدير $-10.04 < \mu_{\text{شركة}1} - \mu_{\text{شركة}2} < 10.14$ يؤكد نفس النتيجة السابقة وهي عدم وجود اختلاف معنوي (ذو أهمية) في معدل أداء الأشخاص في المقابلات الشخصية في الشركتين.

2.2.1.4 الاستدلال للعينات المرتبطة (غير المستقلة) (Inference about Paired Samples)

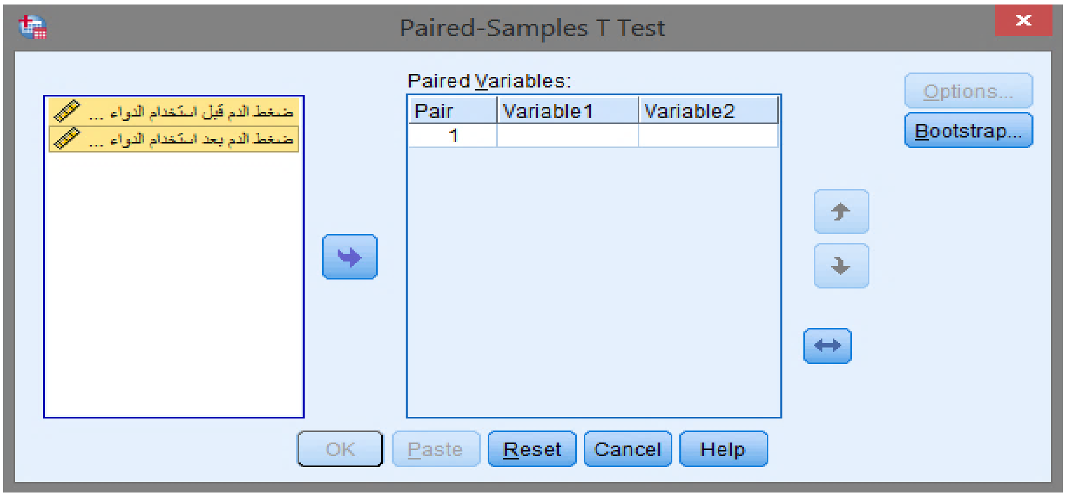
في هذا البند، سنستعرض كيفية تطبيق الاستدلال الإحصائي للفرق بين متوسطي مجتمعين في حال كون العينات غير مستقلة. ولنأخذ المثال التالي لشرح كيفية تنفيذ ذلك في برنامج SPSS؛

البيانات في جدول (7.4)، تمثل قياس ضغط الدم لعينة من 15 مريضا في إحدى المستشفيات قبل وبعد إعطاؤهم دواء تجريبي موسع للأوردة الدموية. والمطلوب هو إجراء استدلال حول وجود فرق معنوي عند المرضى قبل وبعد تناولهم للدواء التجريبي. وهذا الاستدلال سيشمل اختبار الفرضية $H_0 : \mu_D = 0$ ، حيث μ_D هو الوسط الحسابي للفرق بين مشاهدات المجتمعين. قم بإدخال محتويات هذا الجدول في ملف بيانات جديد في SPSS باسم "عينات ضغط الدم"، في عامودين (متغيرين)؛ الأول باسم "عينة1" والثاني باسم "عينة2"، مع إدخال التعريفات "ضغط الدم قبل استخدام الدواء" و"ضغط الدم بعد استخدام الدواء" في خانة (Label).

جدول 7.4: نتائج قياس ضغط الدم لعدد 15 مريض قبل وبعد إعطاؤهم دواء تجريبي.

	العينة 1 (ضغط الدم **قبل** استخدام الدواء)		**العينة 2** (ضغط الدم **بعد** استخدام الدواء)
1	130	1	125
2	128	2	130
3	140	3	135
4	115	4	120
5	146	5	150
6	150	6	152
7	123	7	120
8	160	8	159
9	156	9	160
10	150	10	152
11	112	11	115
12	110	12	115
13	175	13	170
14	160	14	162
15	178	15	180

الآن، في شريط الأوامر العلوي، قم باختيار Analyze>Compare Means>Paired-Samples T Test فتظهر النافذة التالية، (شكل (11.4)). في تلك النافذة، قم أولا بالضغط على زر التحكم (CTRL) في لوحة المفاتيح في جهازك ثم انقر على المتغير الأول ثم الثاني، ("ضغط الدم قبل استخدام الدواء" و"ضغط الدم بعد استخدام الدواء")، وبعد تحديدهما قم بنقلهما معا إلى مربع المتغيرات المرتبطة (Paired Variables)، وستلاحظ أن كلا المتغيرين قد تم وضعهما في نفس الصف.



شكل 11.4: نافذة تحديد العينات المرتبطة للاستدلال حول الفرق بين وسطي مجتمعين لملف بيانات "عينات ضغط الدم".

اضغط موافق بعد ذلك وستظهر النتائج في نافذة المخرجات كما يوضح الجدول (8.4). في الجدول الفرعي الأول (أ)، وكالمعتاد سنلاحظ ظهور النتائج الخاصة بالوسط الحسابي والانحراف المعياري للعينتين، وكذلك الخطأ المعياري لتقدير الوسط الحسابي. أما الجدول الفرعي الثاني (ب)، فيمثل معامل الارتباط بين العينتين والذي سنتناوله بالتفصيل لاحقا عندما نتطرق لموضوع الارتباط بين المتغيرات في الفصل الخامس.

جدول 8.4: نتائج الاستدلال حول الفرق بين الوسطين الحسابيين لعينتين مرتبطتين في ملف البيانات "عينات ضغط الدم".

Paired Samples Statistics (أ)

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 ضغط الدم قبل استخدام الدواء	142.20	15	21.926	5.661
ضغط الدم بعد استخدام الدواء	143.00	15	21.378	5.520

Paired Samples Correlations (ب)

	N	Correlation	Sig.
Pair 1 ضغط الدم قبل استخدام الدواء & ضغط الدم بعد استخدام الدواء	15	.986	**.000**

Paired Samples Test (ج)

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% C. I. of the Difference Lower	Upper			
Pair 1 ضغط الدم قبل استخدام الدواء - ضغط الدم بعد استخدام الدواء	-.800-	3.649	.942	-2.821-	1.221	-.849-	14	**.410**

أما الجدول الفرعي الثالث (ج)، فيضم نتائج الاستدلال حول الفرق بين الوسطين للعينتين المرتبطتين ومنها نستنتج عدم وجود فرق معنوي بين العينتين قبل استخدام الدواء وبعد استخدامه، (P-value =0.410). وتقدير 95% فترة ثقة للفرق بين الوسطين هو $-2.821 < \mu_D < 1.221$ والذي يؤكد نفس النتيجة السابقة.

2.4 اختبار تساوي تباينات المجتمع (Testing the Equality of Population Variances)

تناولنا في الجزء المتعلق بالاستدلال حول الفرق بين وسطي المجتمع، (للعينات المستقلة)، أهمية التحقق من تساوي أو عدم تساوي تباينات المجتمع قبل اختبار تساوي الأوساط. إلا أن التحقق من تساوي تباينات المجتمع، أو بصورة أبسط، التحقق من مدى تجانس المشاهدات بين المتغيرات، عادة ما يكون مفيدا للباحث باختلاف الأسلوب الذي يتبعه في التحليل الإحصائي. وتوجد بضعة اختبارات يمكن استخدامها لاختبار الفرضية؛ $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ من أهمها **اختبار ليفين** (Leven's Test) و**اختبار بارتليت** (Bartlett's Test)، إلا أن برنامج SPSS لا يتضمن اختبار بارتليت لتجانس التباين، لذلك سنتناول اختبار ليفين في البند التالي؛

1.2.4 اختبار ليفن (Leven's Test)

يمكن اعتبار أن اختبار ليفين هو ببساطة تحليل تباين في اتجاه واحد (One-way ANOVA) حيث يكون المتغير الرئيسي فيه (Z) هو الفرق المطلق بين قيمة المشاهدة (Y) ومتوسط المجموعة ($\bar{Y}$) التي تنتمي لها المشاهدة، أي أن $Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|$، (حيث Y_{ij} هي المشاهدة j في المجموعة i). وتكون إحصاءة اختبار ليفن (W) تتبع توزيع F بدرجات حرية (k - 1) و (N - k)، بالصورة التالية:

$$W = \frac{N-k}{k-1} \frac{\sum_{i=1}^{k} N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}$$

حيث؛ k هو عدد المجموعات (المتغيرات)، N_i هو عدد المشاهدات في المجموعة i، N هو عدد المشاهدات الكلي. (وننوه هنا إلى أن صيغة ليفين قد تستخدم أيضا وسيط المجموعة (المتغير) عوضا عن استخدام الوسط الحسابي في الصيغة السابقة).

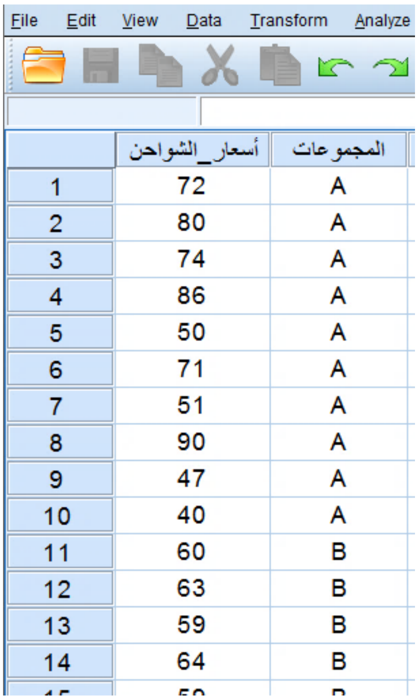
بالنسبة لاختبار ليفين، فإنه يمكن تنفيذه في برنامج SPSS في أكثر من موضع، كما لاحظنا سابقا عند اختبار الفرق بين الأوساط الحسابية، أو كما سنقوم بعرضه من خلال المثال التالي؛

البيانات التالية تمثل أسعار 10 أنواع من شواحن الهواتف النقالة بالدينار الليبي في ثلاثة متاجر مختلفة، A، B، وC، (جدول (9.4))، والمطلوب هنا اختبار ما إذا كان هنالك اختلاف ذو أهمية (معنوية) بين تباين الأسعار في المتاجر، أي اختبار الفرضية $H_0: \sigma_A^2 = \sigma_B^2 = \sigma_C^2$.

جدول 9.4: أسعار شواحن الهواتف النقالة بالدينار الليبي في ثلاثة متاجر مختلفة.

المتجر C	المتجر B	المتجر A	
60	60	72	1
50	63	80	2
51	59	74	3
54	64	86	4
40	59	50	5
45	66	71	6
49	74	51	7
79	60	90	8
59	61	47	9
52	70	40	10

لتنفيذ اختبار ليفين، سنقوم بإدخال البيانات من الجدول (9.4) في ملف بيانات جديد في SPSS باسم "أسعار الشواحن" بحيث يتم إدخال قيم المتجر A أولا في العامود الأول متبوعة بقيم المتجر B في نفس العامود ثم المتجر C. بعد ذلك سنقوم بترميز القيم للمتجر الأول بالرقم 1، وترميز قيم المتجر الثاني بالرقم 2، والمتجر الثالث بالرقم 3، كما نشاهد في الشكل (12.4).

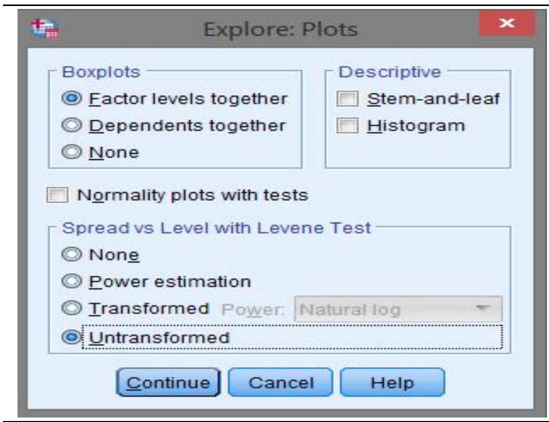


	أسعار_الشواحن	المجموعات
1	72	A
2	80	A
3	74	A
4	86	A
5	50	A
6	71	A
7	51	A
8	90	A
9	47	A
10	40	A
11	60	B
12	63	B
13	59	B
14	64	B

شكل 12.4: ملف بيانات "أسعار الشواحن".

لتطبيق اختبار ليفين، قم من شريط الأوامر العلوي باختيار Analyze>Descriptive Statistics>Explore فتظهر نافذة الاستكشاف (Explore)، (والتي تم التطرق لاستخدامها سابقا في الفصل الثالث). قم في تلك النافذة بنقل المتغير "أسعار_الشواحن" إلى مربع قائمة المتغيرات الأساسية (Dependent List) ونقل المتغير "المجموعات" إلى مربع قائمة متغيرات التصنيف (Factor List).

وحيث أننا مهتمون في هذا المثال باختبار تساوي التباينات فقط، فسنقوم باختيار خيار الرسم البياني (Plots) في الجزء الخاص بالعرض (Display) في أسفل النافذة، لأن اختبار ليفين يقع ضمنها. من جديد قم بالنقر على خيار الرسم (Plots) في أعلى نافذة الاستكشاف على اليمين، فتظهر النافذة الفرعية الخاصة بالرسم (Explore: Plots)، (شكل (13.4)). قم بإلغاء الخيار الافتراضي لرسم الساق والورقة (Stem-and-leaf) لأننا لسنا بحاجة إليه هنا، وفي الجزء السفلي للنافذة (Spread vs Level with Leven Test) قم باختيار الخيار الرابع[1]؛ (بدون تحويل Untransformed) ثم اضغط استمرار للعودة إلى نافذة الاستكشاف الأولى.



شكل 13.4: النافذة الفرعية الخاصة بالرسم لنافذة الاستكشاف.

(يمكنك أيضا، في الشكل (13.4)، ملاحظة إمكانية اختيار عرض الرسم الخاص باختبار الطبيعية (Normality) عن طرق رسم QQ وإجراء اختبارات الطبيعية، وذلك لكل مجموعة). الآن وفي نافذة الاستكشاف قم بالضغط على موافق فتظهر النتائج المطلوبة في نافذة المخرجات. الجدول الأول في نافذة المخرجات كالمعتاد في معظم نتائج برنامج SPSS

[1] الخيارات المتاحة لاختبار ليفين هي خيارات خاصة بتحويل قيم المتغير الأساسي الذي يضم المجموعات.

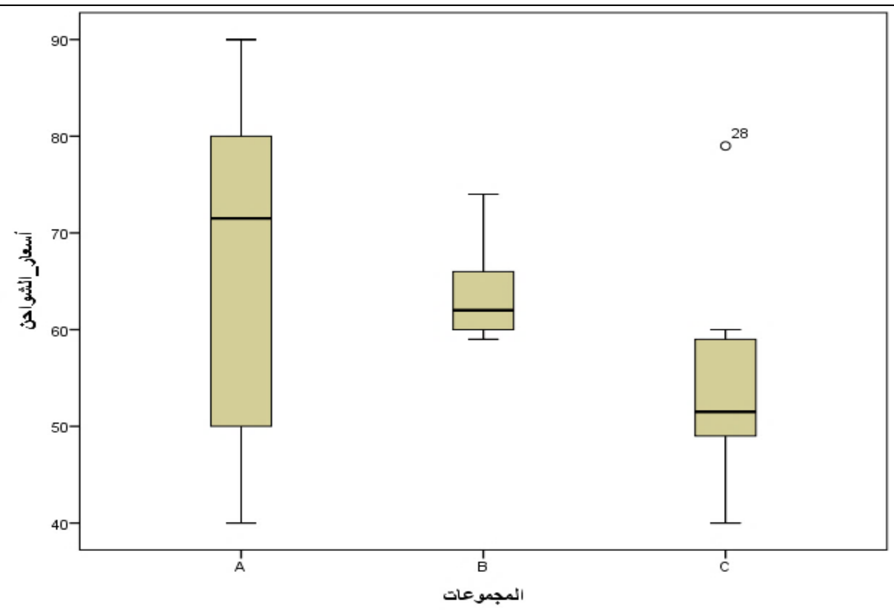
هو خاص بعدد المشـاهدات الموجودة والمفقودة في كل مجموعة ونسـبها. أما الجدول الثاني فيضـم نتائج اختبار ليفين، (جدول (10.4))، وفي هذا الجدول نلاحظ أن العامود الأول إلى اليسار يضم تقسيم النتائج بحسب طريقة حساب إحصاءة ليفين (W) اعتمادا على الوسط الحسابي أو الوسيط أو غيرها من الخيارات.

جدول 10.4: نتائج اختبار ليفين لملف البيانات "أسعار الشواحن".

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
أسعار_الشواحن	Based on Mean	8.810	2	27	**.001**
	Based on Median	4.435	2	27	.022
	Based on Median and with adjusted df	4.435	2	18.862	.026
	Based on trimmed mean	8.608	2	27	.001

من هذه النتائج، (تحديدا P-value = 0.001)، يمكننا القول بأنه يمكن رفض الفرضية الصفرية القائلة بتساوي التباينات في المجتمع، بمعنى أننا نعتقد أسعار الشواحن مختلفة فيما بينها ضمن كل متجر في العموم. وهذه النتيجة يمكن ملاحظتها بصريا من خلال الشكل التالي، (شكل (14.4))، والذي يعرض توزيع المشاهدات لكل مجموعة (متجر) عن طريق شكل الصندوق. ولاحظ أن أسعار الشواحن في المتجر الأول (A) هي الأكثر اختلافا مقارنة بالمتجرين الآخرين. وأما بالنسبة للتمثيل البياني الأخير في نافذة المخرجات، (Spread vs Level Plot)، والذي لن يتم عرضه هنا، فهو يظهر انتشار وسيطات المجموعات الثلاثة حيث تمثل النقطة الأولى (من اليسار) المجموعة B، ثم المجموعة C ثم المجموعة A، وستلاحظ من ذلك الرسم أيضا مدى اختلاف الأسعار في المتجر A عن باقي المتاجر.



شكل 14.4: شكل الصندوق للمجموعات الثلاثة في ملف بيانات "أسعار الشواحن".

ويمكن تنفيذ اختبار ليفين أيضا عند استخدام أسلوب تحليل التباين في اتجاه واحد. في شريط الأدوات العلوي قم باختيار Analyze>Compare Means>One-way ANOVA وفي نافذة التحليل اضغط على الخيارات (Options) واختر منها اختبار تجانس التباين (Homogeneity of variance test) ثم اضغط استمرار. اضغط موافق في النافذة الأصلية فتحصل على النتيجة المطلوبة، (جدول (11.4)).

جدول 11.4: نتائج اختبار ليفين باستخدام أسلوب تحليل التباين، لملف البيانات "أسعار الشواحن".

Test of Homogeneity of Variances

أسعار_الشواحن

Levene Statistic	df1	df2	Sig.
8.810	2	27	**.001**

3.4 جداول الاقتران واختبار الاستقلالية (Contingency Tables and Test for Independency)

جداول الاقتران هي جداول يتم من خلالها تنظيم عرض متغيرين بشكل صفوف وأعمدة، بحيث تمثل الصفوف مستويات (تقسيمات) المتغير الأول وتمثل الأعمدة مستويات المتغير الثاني، وتضم "الخلايا" الناتجة عن التقاء الصفوف والأعمدة التكرارات الناتجة عن "اقتران" أي مستويين. ويتم تكوين جداول الاقتران عادة من المتغيرات الوصفية، إلا أنه يمكن إعادة ترميز المتغيرات الكمية إلى وصفية ثم استخدامها لتكوين الاقتران، وسنبدأ بشرح الحالة الأولى؛

1.3.4 تكوين جداول الاقتران من متغيرات وصفية

(Contingency Tables from Categorical Variables)

لتوضيح طريقة تكوين جداول الاقتران من متغيرين وصفيين لنأخذ المثال التالي؛ البيانات التالية، (في جدول (12.4))، تمثل متغيرين هما مستوى الإنتاج ومستوى الأرق لعينة مكونة من 90 عامل في أحد مصانع الحديد والصلب، حيث تم تقسيم مستوى الأرق إلى ثلاثة مستويات هي منخفض، متوسط، ومرتفع، وتم تقسيم مستوى الإنتاج إلى ثلاثة مستويات أيضا هي ضعيف، مقبول، وجيد.

وكان المطلوب أولا تكوين جدول اقتران لهذين المتغيرين بهدف التعرف على توزيع العمال بحسب مستويات الأرق، وتوزيع مستويات الإنتاج، وتوزيع مستويات الإنتاج مقترنة بمستويات الأرق عند العمال وذلك من خلال التكرارات والنسب لهذه المستويات.

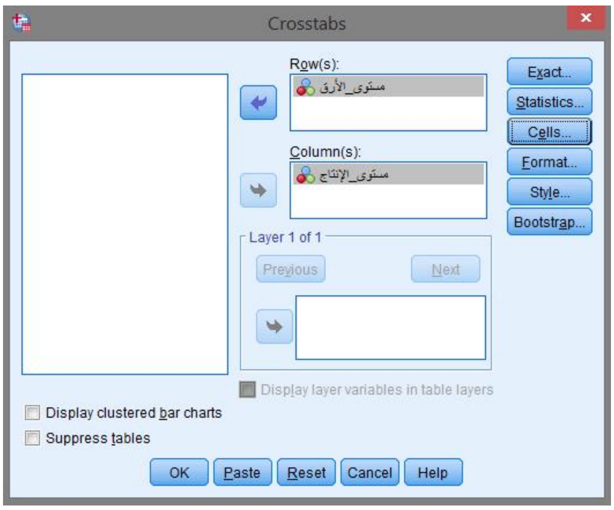
سنبدأ بإدخال هذين المتغيرين في ملف بيانات جديد في SPSS باسم "مستوى الأرق". بحيث نعطي المتغير الأول "مستوى_الإنتاج" القيم: 1، 2، و3 للمستويات: ضعيف، مقبول، وجيد على الترتيب، ونعطي المتغير الثاني

"مستوى_الأرق" القيم: 1، 2، و3 للمستويات: منخفض، متوسط، ومرتفع على الترتيب. مع تعريف هذه القيم بنفس الأسماء في عامود القيم (Values) في نافذة المتغيرات (Variable View).

جدول 12.4: مستوى الإنتاج ومستوى الأرق لعينة من 90 عامل في أحد المصانع.

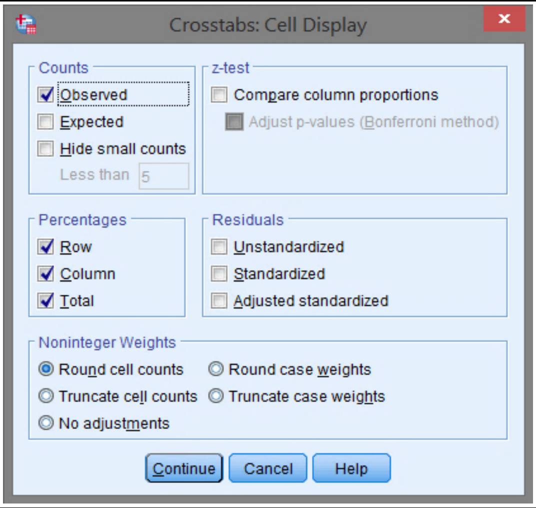
	مستوى الأرق	مستوى الإنتاج		مستوى الأرق	مستوى الإنتاج		مستوى الأرق	مستوى الإنتاج		مستوى الأرق	مستوى الإنتاج		مستوى الأرق	مستوى الإنتاج
1	مرتفع	جيد	21	متوسط	ضعيف	41	مرتفع	ضعيف	61	متوسط	جيد	81	منخفض	جيد
2	مرتفع	ضعيف	22	متوسط	مقبول	42	مرتفع	ضعيف	62	منخفض	مقبول	82	مرتفع	ضعيف
3	مرتفع	ضعيف	23	مرتفع	ضعيف	43	متوسط	ضعيف	63	مرتفع	ضعيف	83	مرتفع	ضعيف
4	منخفض	جيد	24	مرتفع	ضعيف	44	منخفض	مقبول	64	متوسط	جيد	84	منخفض	مقبول
5	مرتفع	ضعيف	25	متوسط	جيد	45	متوسط	مقبول	65	مرتفع	مقبول	85	متوسط	مقبول
6	متوسط	مقبول	26	متوسط	ضعيف	46	منخفض	جيد	66	مرتفع	جيد	86	منخفض	جيد
7	مرتفع	مقبول	27	منخفض	جيد	47	مرتفع	ضعيف	67	منخفض	مقبول	87	متوسط	مقبول
8	متوسط	جيد	28	متوسط	مقبول	48	مرتفع	ضعيف	68	منخفض	مقبول	88	منخفض	جيد
9	مرتفع	جيد	29	منخفض	ضعيف	49	مرتفع	ضعيف	69	متوسط	مقبول	89	متوسط	مقبول
10	مرتفع	جيد	30	مرتفع	جيد	50	مرتفع	ضعيف	70	متوسط	جيد	90	مرتفع	ضعيف
11	مرتفع	ضعيف	31	مرتفع	ضعيف	51	متوسط	جيد	71	مرتفع	ضعيف			
12	متوسط	ضعيف	32	متوسط	ضعيف	52	متوسط	ضعيف	72	مرتفع	مقبول			
13	منخفض	مقبول	33	متوسط	مقبول	53	منخفض	جيد	73	مرتفع	مقبول			
14	متوسط	مقبول	34	مرتفع	ضعيف	54	منخفض	ضعيف	74	منخفض	جيد			
15	منخفض	جيد	35	مرتفع	ضعيف	55	منخفض	مقبول	75	متوسط	مقبول			
16	مرتفع	ضعيف	36	منخفض	جيد	56	منخفض	ضعيف	76	مرتفع	جيد			
17	مرتفع	ضعيف	37	مرتفع	ضعيف	57	منخفض	مقبول	77	مرتفع	ضعيف			
18	متوسط	مقبول	38	متوسط	مقبول	58	منخفض	ضعيف	78	منخفض	جيد			
19	متوسط	ضعيف	39	مرتفع	مقبول	59	منخفض	ضعيف	79	مرتفع	ضعيف			
20	منخفض	جيد	40	متوسط	جيد	60	منخفض	مقبول	80	متوسط	مقبول			

الأن لتكوين جدول الاقتران للمتغيرين، قم في شريط الأدوات العلوي باختيار Analyze>Descriptive Statistics>Crosstabs، فتظهر النافذة الخاصة بتكوين الجداول المشتركة (Crosstabs)، كما نرى في شكل (15.4). في تلك النافذة، قم بنقل المتغير "مستوى_الأرق" إلى مربع الصفوف (Row(s))، ونقل المتغير "مستوى_الإنتاج" إلى مربع الأعمدة (Column(s)).



شكل 15.4: النافذة الخاصة بتكوين جداول الاقتران.

اضـــغط في تلك النافذة على خيار الخلايا (Cells) فتظهر النافذة الفرعية الخاصـــة باختيار ما نود إظهاره في جدول الاقتران المطلوب، (شـــكل (16.4))، قم في تلك النافذة باختيار إظهار النسب للصفوف والأعمدة والمجموع في مربع النسب (Percentages) كما هو موضح في الشكل. ثم اضغط استمرار للعودة للنافذة الســـابقة، (شـــكل (15.4)). وفي تلك النافذة يمكننا إضـــافة تمثل الأعمدة البيانية بحســـب التقســـيمات للمتغيرين، وذلك بـاختيـار عرض الأعمـدة البيـانيـة المصـــفة (Display clustered bar charts) في أســـفل النافذة. قم بعد ذلك بالضـــغط على موافق. ســـتظهر بعد ذلك النتائج في نافذة المخرجات، بحيث يحتوي الجدول الأول على عدد المشـــاهدات الفعلية والقيم المفقودة ونســـبها. أما الجدول الثاني في المخرجات، (جدول (13.4))، فهو جدول الاقتران الخاص بالمتغيرين الوصفيين للمتغيرين "مستوى_الإنتاج" و"مستوى_الأرق".



شكل 16.4: نافذة خيار الخلايا في جداول الاقتران.

جدول 13.4: جدول الاقتران للمتغيرين "مستوى_الإنتاج" و"مستوى_الأرق" في ملف البيانات "مستوى الأرق".

مستوى_الأرق * مستوى_الإنتاج Crosstabulation

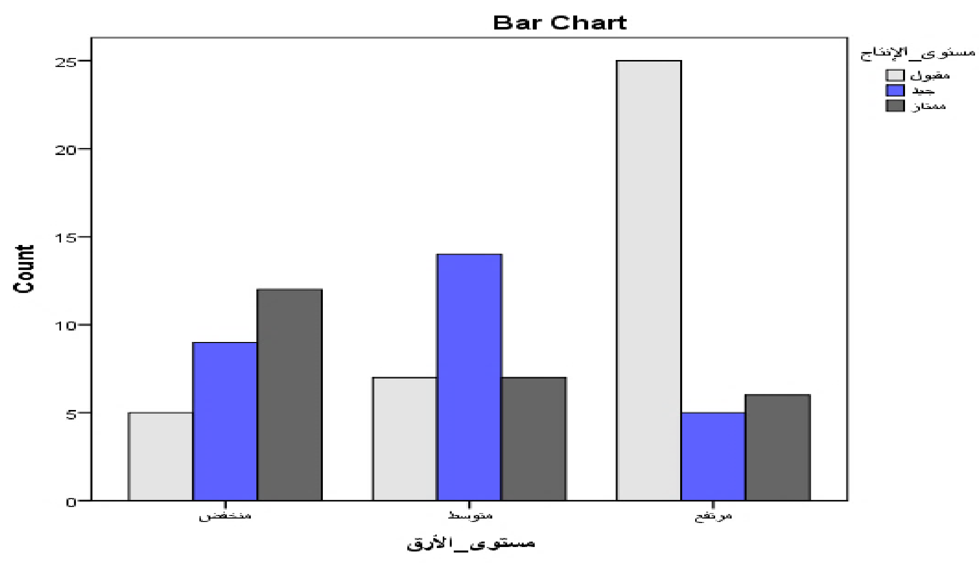
			مستوى_الإنتاج			
			ضعيف	مقبول	جيد	Total
مستوى_الأرق	منخفض	Count	5	9	12	26
		% within مستوى_الأرق	19.2%	34.6%	46.2%	100.0%
		% within مستوى_الإنتاج	13.5%	32.1%	48.0%	28.9%
		% of Total	5.6%	10.0%	13.3%	28.9%
	متوسط	Count	7	14	7	28
		% within مستوى_الأرق	25.0%	50.0%	25.0%	100.0%
		% within مستوى_الإنتاج	18.9%	50.0%	28.0%	31.1%
		% of Total	7.8%	15.6%	7.8%	31.1%
	مرتفع	Count	25	5	6	36
		% within مستوى_الأرق	69.4%	13.9%	16.7%	100.0%
		% within مستوى_الإنتاج	67.6%	17.9%	24.0%	40.0%
		% of Total	27.8%	5.6%	6.7%	40.0%
Total		Count	37	28	25	90
		% within مستوى_الأرق	41.1%	31.1%	27.8%	100.0%
		% within مستوى_الإنتاج	100.0%	100.0%	100.0%	100.0%
		% of Total	41.1%	31.1%	27.8%	100.0%

ونلاحظ وجود أربعة قيم في كل خلية في الجدول. هذه القيم هي التكرار، النسبة للصف، النسبة للعامود، والنسبة للكل. فمثلا في الخلية الأولى؛ (مستوى إنتاج ضعيف ومستوى أرق منخفض)، نجد أن عدد العمال الذين هم إنتاجهم ضعيف ولديهم مستوى أرق منخفض عددهم 5 عمال، ويمثلون 19.2% بالنسبة للعمال الذين لديهم مستوى أرق منخفض، ويمثلون 13.5% بالنسبة للعمال الذين مستوى إنتاجهم ضعيف، ويمثلون 5.6% بالنسبة لكل العمال.

وبالتالي، يمكننا الحصول على العديد من المعلومات من جدول الاقتران عن توزيع العمال بحسب مستويات الإنتاج ومستويات الأرق أهمها ما يلي:

- الخلية التي لها أعلى نسبة للمجموع هي التي تمثل العمال الذين مستوى إنتاجهم ضعيف ومستوى الأرق لديهم مرتفع (27.8%)، وهذا يعني أن العمال الذين يعانون من الأرق في النوم بصورة كبيرة يكون مستوى إنتاجهم ضعيف. ويؤيد تلك النتيجة أيضا نسبة المجموع للعمال الذين مستوى إنتاجهم جيد ومستوى الأرق لديهم منخفض (13.3%).
- نسبة العمال الذين لديهم مستوى أرق مرتفع هي 40% وهي نسبة تُعد كبيرة، أما نسبة العمال الذين لديهم مستوى أرق متوسط فهي 31.1%، وبالتالي فإنه بجمع هتين النسبتين يمكن القول بأن 71.1% من العمال يعانون من مستويات أرق متوسطة إلى مرتفعة في النوم.
- نسبة العمال الذين إنتاجهم جيد هو 27.8% أي أقل من ثلث العمال، ونسبة العمال الذين إنتاجهم ضعيف هو 41.1% وهذا يعكس الوضع السيئ للإنتاج في هذا المصنع.

أما النتيجة الأخيرة في نافذة المخرجات، (شكل (17.4))، فهي الأعمدة البيانية المصنفة بحسب مستويات الإنتاج ومستويات الأرق لدى العمال. وهذه الأعمدة تمثل النتائج المذكورة سابقا بشكل مرئي.



شكل 17.4: الأعمدة البيانية المصنفة لملف البيانات "مستوى الأرق".

2.3.4 تكوين جداول الاقتران من متغيرات كمية (Contingency Tables from Quantitative Variables)

في بعض الدراسات، قد نقوم بتحويل بعض المتغيرات الكمية إلى وصفية، أي إلى مستويات، وذلك بهدف استخدامها في بعض الأساليب الإحصائية التي تتطلب هذا النوع من المتغيرات، أو بهدف الحصول على تلخيص لنمط معين في البيانات. ولتكوين جدول اقتران من متغير كمي وآخر وصفي أو من متغيران كميان، لنأخذ المثال التالي؛

البيانات في الجدول (14.4) تمثل معدلات فيتامين دال (Vit. D) وأعمار عينة مكونة من 60 شخصا إضافة لتصنيفهم ذكور وإناث.

جدول 14.4: معدلات فيتامين دال (Vit. D) وأعمار عينة مكونة من 60 شخصا.

النوع	معدل فيتامين دال	العمر		النوع	معدل فيتامين دال	العمر		النوع	معدل فيتامين دال	العمر	
ذكر	26	72	41	أنثى	26	55	21	ذكر	26	53	1
أنثى	81	28	42	أنثى	29	69	22	ذكر	28	61	2
ذكر	66	39	43	أنثى	20	68	23	ذكر	20	81	3
أنثى	20	27	44	ذكر	36	58	24	أنثى	51	25	4
أنثى	24	48	45	أنثى	25	57	25	أنثى	30	20	5
أنثى	29	67	46	ذكر	79	24	26	أنثى	25	71	6
أنثى	20	25	47	أنثى	20	20	27	أنثى	21	80	7
ذكر	78	39	48	أنثى	20	90	28	ذكر	28	65	8
أنثى	23	46	49	أنثى	25	26	29	ذكر	74	30	9
أنثى	26	58	50	أنثى	34	35	30	ذكر	80	22	10
ذكر	80	67	51	أنثى	39	38	31	ذكر	26	77	11
ذكر	28	81	52	ذكر	25	56	32	أنثى	30	50	12
ذكر	45	36	53	ذكر	75	47	33	ذكر	24	65	13
ذكر	69	35	54	أنثى	26	30	34	أنثى	22	23	14
أنثى	20	24	55	ذكر	77	25	35	ذكر	26	61	15
ذكر	84	31	56	أنثى	29	68	36	أنثى	24	46	16
ذكر	80	21	57	ذكر	79	35	37	ذكر	76	31	17
ذكر	32	84	58	ذكر	80	26	38	ذكر	78	35	18
ذكر	75	78	59	أنثى	24	24	39	ذكر	80	26	19
ذكر	77	88	60	أنثى	36	39	40	أنثى	23	48	20

ونلاحظ من هذه البيانات أن المتغيران "معدل فيتامين دال"، (المقاس بوحدة نانومول) و"العمر"، (المقاس بالسنة) هما متغيران كميان، ولتكوين جدول اقتران خاص بهما يجب أولا تحويلهما إلى متغيرات وصفية تأخذ مستويات محددة.

لنقم أولا بإدخال المتغيرات الثلاثة من جدول (14.4) في ملف بيانات جديد في SPSS باسم "فيتامين دال"، (شكل (18.4))، بحيث يتم إعطاء القيمة 1 للذكور والقيمة 2 للإناث، ثم تعريف تلك القيم في خانة القيم (Values) كالمعتاد، وسنعطي المتغيرات الثلاثة الأسماء "العمر"، "فيتامين_دال"، و"النوع".

	العمر	فيتامين_دال	النوع
1	53	26	1
2	61	28	1
3	81	20	1
4	25	51	2
5	20	30	2
6	71	25	2
7	80	21	2
8	65	28	1
9	30	74	1
10	22	80	1
11	77	26	1
12	50	30	2
13	65	24	1

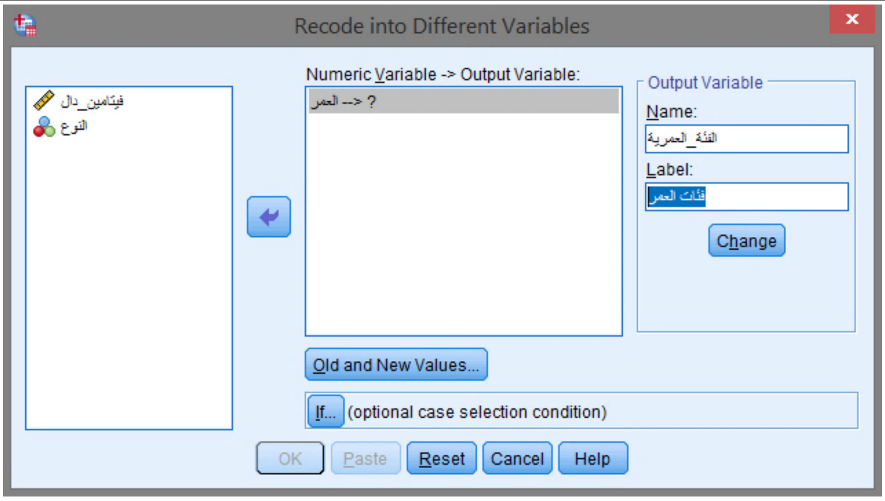
شكل 18.4: المتغيرات "العمر"، "فيتامين_دال"، و"النوع" في ملف البيانات "فيتامين دال".

في الفصل الثاني، وفي الجزء الخاص بإعادة ترميز القيم في متغير مختلف (Recode into Different Variable)، تناولنا كيفية تكوين متغير جديد عن طريق تعريف قيم أو مستويات جديدة له من المتغير الأصلي، وهنا سنقوم أيضا بتعريف متغيران جديدان؛

بالنسبة للمتغير الأول، "العمر"، سنقوم بتحويله إلى متغير يأخذ ثلاثة مستويات هي الفئات العمرية؛ (أقل من 30)، (من 30 إلى 50)، و(51 فأكثر)، وسنستخدم القيم 1، 2، و3 لترميز الفئات العمرية السابقة على التوالي.

وأما المتغير "فيتامين_دال" فسيتم إعطائه المستويات الثلاثة المُعدة من قِبل الأطباء المتخصصين وهي (أقل من 30) وفيها يصنف الشخص بأنه يعاني من نقص في فيتامين دال، و(من 30 إلى 75) ويصنف فيها الشخص بأن مستوى فيتامين دال غير كافي، و(أكثر من 75) وهو المستوى الطبيعي لفيتامين دال عند الإنسان البالغ. وسنعطي القيمة 1 للمستوى (أقل من 30) والقيمة 2 للمستوى (من 30 إلى 75) والقيمة 3 للمستوى (أكثر من 75).

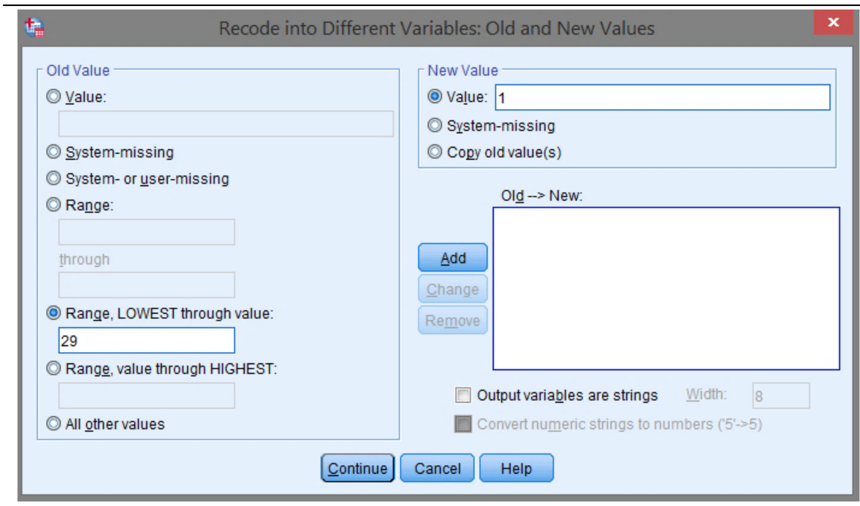
الآن قم باختيار أمر تحويل (Transform) من شريط الأوامر العلوي، ثم اختر إعادة ترميز في متغير مختلف (Recode into Different Variables). ستظهر عندها نافذة إعادة الترميز في متغير جديد، كما يُشاهد في الشكل (19.4). سنقوم باختيار المتغير "العمر"، ونقله إلى مربع المتغيرات (Numeric Variable). وفي خانة المتغير المُخرج (Output Variable) إلى اليمين، لنقم بتسمية المتغير الجديد بالاسم "الفئة_العمرية" وكتابة "فئات العمر" في مربع الوصف (Label)، ثم نضغط على أمر تغيير (Change)، فيلاحظ حدوث هذا التغيير في مربع المتغيرات حيث سيظهر فيها النص "العمر --> الفئة_العمرية".



شكل 19.4: نافذة إعادة الترميز في متغير مختلف لملف البيانات "فيتامين دال".

بعد ذلك سنقوم بتعريف القيم التي نود إعادة كتابتها في المتغير الجديد باستخدام قيم متغير "العمر"، فنقوم باختيار أمر تعريف القيم القديمة والجديدة (Old and New Values) في الشكل (19.4)، فتظهر نافذة جديدة كما يُرى في الشكل (20.4).

في هذه النافذة، لتعيين القيمة الأولى، والتي تشمل كل الأعمار التي هي أقل من 30 سنة، نحتاج لاستخدام الخيار المدى، القيمة فأقل (Range, LOWEST through value). فنكتب أولا القيمة المحددة للتصنيف الأول وهي القيمة 1 في خانة القيمة الجديدة (New Value)، ثم نكتب القيمة 29، (وهي القيمة[1] الأقل من 30 سنة مباشرة في متغير العمر)، ثم نضغط إضافة (Add).



شكل 20.4: نافذة تعريف القيم القديمة والجديدة لمتغير "العمر" في خيار إعادة الترميز في متغير مختلف لملف البيانات "فيتامين دال".

[1] ننوه هنا إلى أنه إذا ما تم كتابة القيمة 30 في خانة (Range, LOWEST through value) فإن هذا التصنيف سيشمل على العمر 30 سنة من ضمنه.

لتعيين القيمة الثانية، وهي القيمة 2 للتصنيف الثاني وهو الأعمار من 30 إلى 50 سنة، نكتب القيمة 2 في خانة القيمة الجديدة (New Value) ثم نضغط على الخيار المدى (Range) فتظهر خانتان فارغتان أسفل منه. يتم إدخال القيمة 30 في الخانة العليا والقيمة 50 في الخانة السفلى والضغط على (Add).

ولتعيين القيمة الثالثة، وهي القيمة 3 للتصنيف الثالث، وتمثل الأشخاص الذين أعمارهم 51 سنة فأكثر، نكتب القيمة 3 في خانة القيمة الجديدة (New Value) ثم نضغط على الخيار المدى، القيمة فأكبر من (Range, value through HIGHEST) ونكتب فيه القيمة 51، ثم نضغط على (Add). وبعد الانتهاء من تعيين كل القيم المحددة للتصنيفات السابقة نضغط على استمرار (Continue). سنعود بعد ذلك للنافذة السابقة، (شكل (19.4))، نقوم بعدها بالضغط على موافق فيظهر المتغير الجديد "الفئة_العمرية" في نافذة عرض البيانات في ملف البيانات "فيتامين دال".

الآن وللانتهاء من تعريف المتغير الجديد "الفئة_العمرية"، قم بتعريف القيم 1، 2، و3 في عامود القيم (Values) في نافذة المتغيرات (Variable View) للمستويات الثلاثة؛ (أقل من 30)، (من 30 إلى 50)، و(51 فأكثر).

من جديد، وبنفس الكيفية، سنقوم بتعريف متغير جديد لتحويل متغير "فيتامين_دال" إلى متغير وصفي ذو ثلاثة مستويات. في نافذة إعادة ترميز في متغير مختلف[1] (Recode into Different Variables) سنقوم باختيار المتغير "فيتامين_دال"، ونقله إلى مربع المتغيرات (Numeric Variable). وفي خانة المتغير المُخرج (Output Variable) إلى اليمين، نقوم بتسمية المتغير الجديد بالاسم "مستويات_فيتامين_دال" وكتابة "مستويات فيتامين دال" في مربع الوصف (Label)، ثم نضغط على أمر تغيير (Change)، فنلاحظ حدوث هذا التغيير في مربع المتغيرات حيث سيظهر فيها النص "فيتامين_دال --> مستويات_فيتامين_دال".

بعد ذلك نقوم بتعريف قيم المتغير الجديد فنقوم باختيار أمر تعريف القيم القديمة والجديدة (Old and New Values) ثم تعيين القيمة الأولى وهي 1، والتي تشمل كل معدلات فيتامين دال التي هي أقل من 30 نانومول، في الخيار المدى، القيمة فأقل (Range, LOWEST through value)، ونكتب القيمة 1 في خانة القيمة الجديدة (New Value)، ثم نكتب القيمة 29، (وهي القيمة الأقل من 30 نانومول مباشرة)، ثم نضغط إضافة (Add).

[1] عند استخدام الأوامر في برنامج SPSS بصورة متتالية في نفس الجلسة لملف بيانات ما، فإنك ستلاحظ وجود الخيارات والإعدادات التي قمت بها موجودة كما هي، وهذه الخاصية قد تكون مفيدة في كثير من الأحيان عندما تقوم بتكرار تحليل إحصائي معين وإجراء تغييرات محددة، أما في حالات أخرى، مثل المثال الحالي، فإننا نكون بحاجة لإلغاء ما تم تعريفه للمتغير السابق وكتابة التعريفات الجديدة.

لتعيين القيمة الثانية، وهي القيمة 2 للتصنيف الثاني وهي المعدلات من 30 إلى 75 نانومول، نكتب القيمة 2 في خانة القيمة الجديدة (New Value) ثم نضغط على الخيار المدى (Range) فتظهر خانتان فارغتان أسفل منه. يتم إدخال القيمة 30 في الخانة العليا والقيمة 75 في الخانة السفلى والضغط على (Add).

ولتعيين القيمة الثالثة، وهي القيمة 3 للتصنيف الثالث، وهي معدلات فيتامين دال التي هي أكثر من 75 نانومول، نكتب القيمة 3 في خانة القيمة الجديدة (New Value) ثم نضغط على الخيار المدى، القيمة فأكبر من (Range, value through HIGHEST) ونكتب فيه القيمة 76، ثم نضغط على (Add). وبعد الانتهاء من تعيين كل القيم المحددة للتصنيفات السابقة نضغط على استمرار (Continue). سنعود بعد ذلك للنافذة السابقة، (شكل (19.4))، نقوم بعدها بالضغط على موافق فيظهر المتغير الجديد "مستويات_فيتامين_دال" في نافذة عرض البيانات في ملف البيانات "فيتامين دال"، وللانتهاء من تعريف المتغير الجديد "مستويات_فيتامين_دال"، قم بتعريف القيم 1، 2، و3 في عامود القيم (Values) في نافذة المتغيرات (Variable View) للمستويات الثلاثة؛ (أقل من 30)، (من 30 إلى 75) و(أكثر من 75)، ويكون الشكل النهائي لملف البيانات "فيتامين دال"، شكل (21.4).

	العمر	فيتامين_دال	النوع	الفئة_العمرية	مستويات_فيتامين_دال
1	53	26	ذكر	51 فأكثر	أقل من 30
2	61	28	ذكر	51 فأكثر	أقل من 30
3	81	20	ذكر	51 فأكثر	أقل من 30
4	25	51	أنثى	أقل من 30	من 30 إلى 75
5	20	30	أنثى	أقل من 30	من 30 إلى 75
6	71	25	أنثى	51 فأكثر	أقل من 30
7	80	21	أنثى	51 فأكثر	أقل من 30
8	65	28	ذكر	51 فأكثر	أقل من 30
9	30	74	ذكر	من 30 إلى 50	من 30 إلى 75
10	22	80	ذكر	أقل من 30	أكثر من 75
11	77	26	ذكر	51 فأكثر	أقل من 30
12	50	30	أنثى	من 30 إلى 50	من 30 إلى 75
13	65	24	ذكر	51 فأكثر	أقل من 30

شكل 21.4: ملف البيانات "فيتامين دال" بعد تعريف المتغيرين الجديدين "الفئة_العمرية" و"مستويات_فيتامين_دال".

بعد أن تم تحويل المتغيران الكميان "العمر" و"فيتامين_دال" في ملف البيانات "فيتامين دال" إلى متغيرين وصفيين رتبيين، سنقوم الآن بتكوين جدول الاقتران لهما؛

قم في شريط الأدوات العلوي باختيار Analyze>Descriptive Statistics>Crosstabs، فتظهر النافذة الخاصة بتكوين الجداول المشتركة (Crosstabs). في تلك النافذة، قم بنقل المتغير "الفئة العمرية" إلى مربع الصفوف (Row(s))، ونقل المتغير "مستويات_فيتامين_دال " إلى مربع الأعمدة (Column(s)).

اضغط في تلك النافذة على خيار الخلايا (Cells) وقم باختيار إظهار النسب للصفوف والأعمدة والمجموع في مربع النسب (Percentages)، ثم اضغط استمرار للعودة للنافذة السابقة.

وفي تلك النافذة قم بإضافة تمثيل الأعمدة البيانية بحسب التقسيمات للمتغيرين، وذلك باختيار عرض الأعمدة البيانية المصنفة (Display clustered bar charts) في أسفل النافذة. قم بعد ذلك بالضغط على موافق.

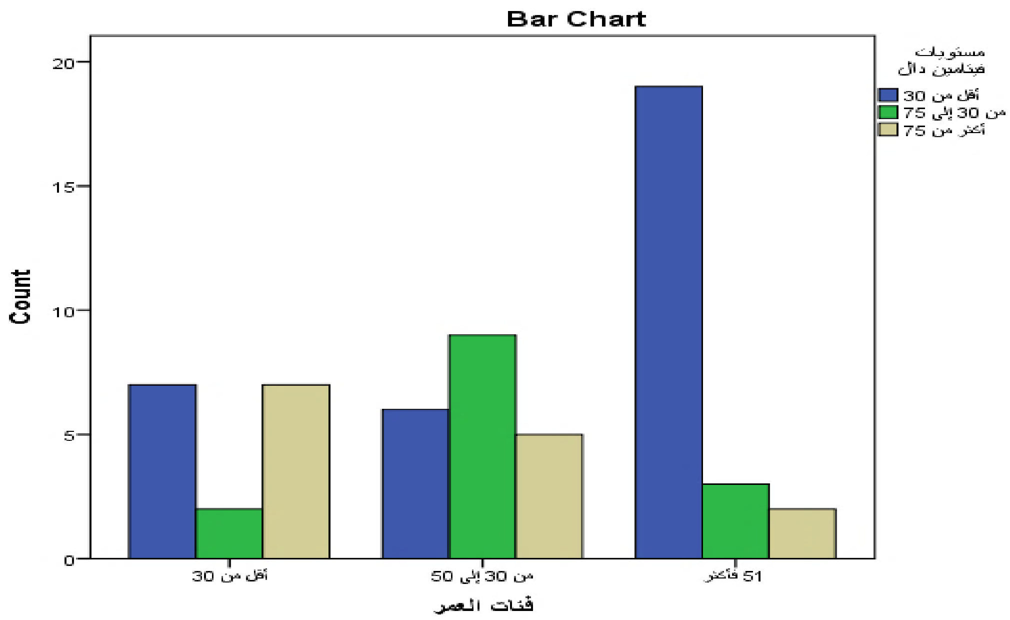
ستظهر بعد ذلك النتائج في نافذة المخرجات، بحيث يحتوي الجدول الأول على عدد المشاهدات الفعلية والقيم المفقودة ونسبها. أما الجدول الثاني فهو جدول الاقتران، والذي يظهر في جدول (15.4)، ومنه نستطيع ملاحظة أن النسبة الأكثر بروزا هي نسبة الأشخاص الأكبر سنا (15 سنة فأكثر) الذين لديهم نقص حاد في معدل فيتامين دال، (31.7%).

جدول 15.4: جدول الاقتران للمتغيرين "الفئة_العمرية" و"مستويات_فيتامين_دال" في ملف البيانات "فيتامين دال".

فئات العمر * مستويات فيتامين دال Crosstabulation

			مستويات فيتامين دال			Total
			أقل من 30	**من 30 إلى 75**	**أكثر من 75**	
فئات العمر	**أقل من 30**	Count	7	2	7	16
		% within فئات العمر	43.8%	12.5%	43.8%	100.0%
		% within مستويات فيتامين دال	21.9%	14.3%	50.0%	26.7%
		% of Total	11.7%	3.3%	11.7%	26.7%
	من 30 إلى 50	Count	6	9	5	20
		% within فئات العمر	30.0%	45.0%	25.0%	100.0%
		% within مستويات فيتامين دال	18.8%	64.3%	35.7%	33.3%
		% of Total	10.0%	15.0%	8.3%	33.3%
	51 فأكثر	Count	19	3	2	24
		% within فئات العمر	79.2%	12.5%	8.3%	100.0%
		% within مستويات فيتامين دال	59.4%	21.4%	14.3%	40.0%
		% of Total	31.7%	5.0%	3.3%	40.0%
Total		Count	32	14	14	60
		% within فئات العمر	53.3%	23.3%	23.3%	100.0%
		% within مستويات فيتامين دال	100.0%	100.0%	100.0%	100.0%
		% of Total	53.3%	23.3%	23.3%	100.0%

كما يمكن، من الأعمدة البيانية المصنفة للمتغيرين (شكل (22.4))، ملاحظة ارتفاع هذه النسبة مقارنة بالنسب أو التكرارات الأخرى.



شكل 22.4: الأعمدة البيانية المصنفة للمتغيرين "الفئة_العمرية" و"مستويات_فيتامين_دال" في ملف البيانات "فيتامين دال".

3.3.4 اختبار الاستقلالية (Test for Independency)

يُعد اختبار الاستقلالية من الاختبارات الهامة المستخدمة في دراسة العلاقات بين المتغيرات الوصفية، باستخدام جداول الاقتران، ويُعرف أيضا باختبار **بيرسون مربع كاي** (Pearson's Chi-Square Test). وتُعرف إحصاءة الاختبار الخاصة به، والتي تتبع توزيع مربع كاي بدرجات حرية تساوي ($(r\text{-}1)(c\text{-}1)$)، بالصورة؛

$$\chi_c^2 = \sum_{i=1}^{rc} \frac{(o_i - e_i)^2}{e_i}$$

حيث o_i هو التكرار المشاهد في الخلية، $e_i = \frac{\text{مجموع العامود} \times \text{مجموع الصف}}{\text{المجموع الكلي}}$ هو التكرار المتوقع للخلية، r=عدد الصفوف، و c=عدد الأعمدة.

ويتم في اختبار الاستقلالية اختبار الفرضية الصفرية؛ **توجد استقلالية بين المتغيرين (الصفوف والأعمدة)** :H_0، وبالتالي فإن رفض الفرضية الصفرية يعني وجود علاقة ذات معنوية بين بعض أو كل مستويات المتغيرين في جدول الاقتران.

وقبل البدء في تنفيذ اختبار الاستقلالية، سنعرف بعض المقاييس الإحصائية التي تستخدم كمقاييس إضافية لدراسة الارتباط بين متغيري جدول الاقتران، والتي تعتمد على قيمة إحصاء مربع كاي في حسابها، وتسمى **مقاييس الترابط بناء على مربع كاي** (Measures of Association based on Chi-Square)؛

1.3.3.4 معامل فاي (Phi Coefficient)

يعرف معامل فاي للارتباط بالصورة التالية:

$$\varphi = \sqrt{\frac{\chi_c^2}{n}}$$

حيث χ_c^2 هي قيمة إحصاءة مربع كاي المحسوبة من جدول الاقتران، و n هو عدد المشاهدات (التكرارات) الكلي في جدول الاقتران. وكلما اقتربت قيمة φ من الصفر كلما دل ذلك على ضعف العلاقة بين المتغيرين الوصفيين، وبالمقابل كلما اقتربت من الواحد الصحيح كلما دل ذلك على قوة العلاقة.

2.3.3.4 معامل الاقتران (Contingency Coefficient)

معامل الاقتران، والذي يُعرف أيضا **بمعامل الاقتران لبيرسون** (Pearson's Contingency Coefficient)، هو أيضا أحد المقاييس التي تستخدم لقياس درجة العلاقة بين متغيري جدول الاقتران، ويعتمد أيضا مثل معامل فاي على إحصاءة مربع كاي، ويُعرف بالصورة:

$$C = \sqrt{\frac{\chi_c^2}{n + \chi_c^2}}$$

حيث χ_c^2 هي قيمة إحصاءة مربع كاي ، و n هو عدد التكرارات الكلي في جدول الاقتران. وينطبق عليه ما ينطبق على معامل كاي في التعليق على قيمته المحسوبة. ويمكننا أيضا ملاحظة العلاقة بين صيغتي كل من المعاملين كالتالي:

$$C = \sqrt{\frac{\varphi^2}{1 + \varphi^2}}$$

3.3.3.4 معامل كرامر *V* (Cramer's *V* Coefficient)

يُعرف معامل كرامر V، والذي يعتمد هو الآخر في حسابه على إحصاءة مربع كاي، بالصورة التالية:

$$V = \sqrt{\frac{\chi_c^2}{nk}}$$

حيث χ_c^2 هي قيمة إحصاءة مربع كاي، n هو عدد التكرارات الكلي في جدول الاقتران، و k هو الحد الأدنى لعدد الصفوف وعدد الأعمدة في جدول الاقتران، أي أن $k = \text{Min.}(r,c)$. ويكون التعليق على قيمته كما هو الحال مع

المعاملين السابقين، معامل فاي ومعامل الاقتران. ويمكن ملاحظة العلاقة بين صيغتي كل من معامل كرامر ومعامل فاي كالتالي:

$$V = \sqrt{\frac{\varphi^2}{k}}$$

4.3.3.4 معامل كندل تاو (Kendall's Tau Coefficient)

وهو معامل يُستخدم أيضا لحساب الارتباط بين متغيرين وصفيين، إلا أنه يتميز تحديدا بالتعامل مع المتغيرات الوصفية الرتبية. وتتوفر لهذا المعامل ثلاثة صيغ بحسب طبيعة جدول الاقتران سنتناولها فيما يلي:

- **معامل كندل تاو B** (Kendall's Tau-B)

ويستخدم عادة في جداول الاقتران المربعة، (التي تتساوى فيها عدد مستويات الصفوف والأعمدة). وتُعرف صيغة المعامل بالصورة:

$$\tau_b = \frac{P - Q}{\frac{n(n-1)}{2}}$$

حيث n هو عدد التكرارات الكلي في جدول الاقتران، و P و Q هما عدد الأزواج (من الخلايا) المتوافقة (Concordant) وغير المتوافقة (Discordant) على الترتيب. ولتوضيح كيفية حساب هاتين القيمتين من جداول الاقتران ذات الحجم 2×2، أي التي تحتوي على صفين وعامودين، (حيث أن حساب القيمتين من الجداول الأكبر حجما يكون أكثر تعقيدا باستخدام الحل اليدوي)، اعتبر جدول الاقتران ذو القيم الافتراضية التالي، (جدول (16.4))؛

جدول 16.4: جدول اقتران 2×2 افتراضي.

		المتغير الأول	
		المستوى 1	المستوى 2
المتغير الثاني	المستوى 1	***a***	***b***
	المستوى 2	***c***	***d***

لهذا الجدول، يتم حساب قيم كل من P و Q بالصورة التالية:

$$Q = a \times d \text{ و } P = b \times c$$

- **معامل كندل تاو B مع التصحيح في حالة التساوي** (Kendall's Tau-B with correction for ties)

وهو نفس المعامل السابق مع إضافة تصحيح في الحسابات للخلايا عند تساوي قيم P و Q.

- **معامل كندل تاو C** (Kendall's Tau-C)

ويستخدم هذا المعامل في جداول الاقتران غير المربعة، أي التي يكون فيها عدد الصفوف وعدد الأعمدة غير متساوي، ويأخذ الصيغة التالية:

$$\tau_c = \frac{P - Q}{n^2(\min(r, c) - 1)2\min(r, c)}$$

حيث $\min(r, c)$ هي القيمة الأقل للصفوف والأعمدة.

وكما هو الحال مع معاملات الارتباط السابقة، فإن القيمة الأعلى لمعامل كندل، أي الأقرب للواحد الصحيح، تكون مؤشرا على قوة العلاقة بين المتغيرين، ونظرا لوجود امكانية أن تكون $P < Q$ فإن معامل كندل قد يأخذ القيمة السالبة، وهذا يكون مؤشرا على وجود علاقة عكسية بين المتغيرين.

5.3.3.4 معامل جاما (Gamma Coefficient)

يعتمد معامل جاما للارتباط بين متغيري جدول الاقتران على قيم P و Q أيضا مثل معامل كندل، وتُعرف صيغته كالتالي:

$$\gamma = \frac{P - Q}{P + Q}$$

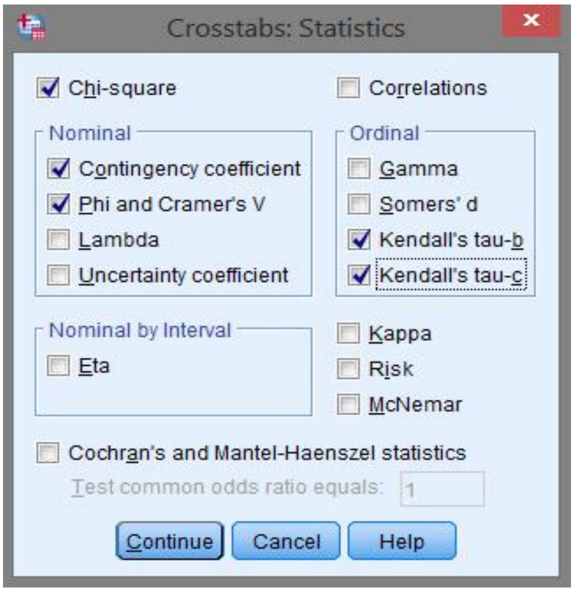
وتكون قيمته محصورة بين -1 و 1 كما هو الحال مع معامل كندل.

الآن سنقوم بشرح خطوات تنفيذ اختبار الاستقلالية لجداول الاقتران مع حساب معاملات الارتباط السابقة، ولنستخدم ملف البيانات الأخير؛ "فيتامين دال". في هذا الملف، سنبدأ باختبار العلاقة بين المتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية".

في شريط الأدوات العلوي قم باختيار Analyze>Descriptive Statistics>Crosstabs، فتظهر النافذة الخاصة بتكوين الجداول المشتركة (Crosstabs). في تلك النافذة، قم بنقل المتغير "الفئة العمرية" إلى مربع الصفوف (Row(s))، ونقل المتغير "مستويات_فيتامين_دال " إلى مربع الأعمدة (Column(s)).

الآن قم بالضغط على خيار الإحصاءات (Statistics) في يمين النافذة، فتظهر نافذة جديدة، (شكل (23.4))، خاصة باختيار اختبار مربع كاي للاستقلالية ومعاملات الارتباط المتعلقة بجداول الاقتران. ولاحظ أننا قمنا في تلك النافذة باختيار اختبار مربع كاي، ومعامل الاقتران ومعاملي فاي وكرامر V ومعاملي كندل تاو B و C.

نضغط بعد تلك الاختيارات استمرار (Continue) للعودة للنافذة الأصلية. ولاحظ أننا لن نقوم باختيار إظهار نسب الخلايا ونسب مجاميع الصفوف والأعمدة من خيار الخلايا (Cells) لأنه قد سبق التطرق لها. الآن اضغط موافق في النافذة الأصلية للحصول على النتائج المطلوبة.



شكل 23.4: نافذة اختيار اختبار الاستقلالية ومعاملات الارتباط لجداول الاقتران.

ستظهر في نافذة المخرجات أربعة جداول؛ الجدول الأول يمثل عدد المشاهدات الفعلية والقيم المفقودة ونسبها، والجدول الثاني هو جدول الاقتران للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية"، والجدول الثالث، (جدول (17.4))، يضم نتائج اختبار مربع كاي للاستقلالية، ولاحظ من الصف الأول من الجدول أن $\chi_c^2 = 16.552$ وأن القيمة الاحتمالية (P-value = 0.002) وهذا يدفعنا لرفض الفرضية الصفرية القائلة بوجود استقلالية بين المتغيرين (الصفوف والأعمدة)، والاستدلال بوجود علاقة ذات أهمية بين معدلات فيتامين دال وأعمار الأشخاص في العموم. وبربط هذه النتيجة مع ما توصلنا له من جدول (15.4) وشكل (22.4)، نستطيع تأكيد انخفاض معدلات فيتامين دال لدى الأشخاص في الفئة العمرية 51 سنة فأكبر.

جدول 17.4: نتيجة اختبار مربع كاي للاستقلالية للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية" لملف البيانات "فيتامين دال".

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	16.552[a]	4	**.002**
Likelihood Ratio	16.285	4	.003
Linear-by-Linear Association	7.956	1	.005
N of Valid Cases	60		

a. 4 cells (44.4%) have expected count less than 5. The minimum expected count is 3.73.

والملاحظة الموجودة في أسفل الجدول هي للتنبيه بوجود 4 تكرارات متوقعة (e_i) قيمها أقل من 5 في حساب قيمة إحصاءة مربع كاي.

أما الجدول الرابع في نافذة المخرجات، (جدول (18.4))، فيضم النتائج الخاصة بقيم معاملات الارتباط الخاصة بجدول الاقتران للمتغيرين المعنيين. وبالنظر إلى قيم الاحتمالية (P-value) لكل المعاملات، نجد أن جميعها أقل من 0.05 وبالتالي نستطيع القول بوجود علاقة ذات معنوية عالية بين معدلات فيتامين دال والفئة العمرية. ولاحظ ارتفاع قيم بعض المعاملات مثل معامل فاي ومعامل الاقتران، ولاحظ أيضا الإشارة السالبة لمعاملات كندل مما يدل على وجود علاقة عكسية بين المتغيرين.

جدول 18.4: معاملات الارتباط لجدول الاقتران للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية"، للملف "فيتامين دال".

Symmetric Measures

		Value	Asymptotic Standardized Error[a]	Approximate T[b]	Approximate Significance
Nominal by Nominal	Phi	.525			**.002**
	Cramer's V	.371			**.002**
	Contingency Coefficient	.465			**.002**
Ordinal by Ordinal	Kendall's tau-b	-.347	.114	-3.020	**.003**
	Kendall's tau-c	-.329	.109	-3.020	**.003**
N of Valid Cases		60			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

4.3.4 اختبار الاستقلالية باستخدام متغيرات تصنيف إضافية

(Test for Independency using additional Classification Variables)

في بعض الدراسات، قد نكون بحاجة لتحليل العلاقة بين متغيرين وصفيين بوجود متغير ثالث أو أكثر يتضمن/تتضمن مستويات يمكن تصنيف العلاقة بين المتغيرين الأساسيين بناء عليها، ولتوضيح الفكرة لنعد لملف البيانات "فيتامين دال" مرة أخرى، حيث أننا سنقوم الآن باختبار العلاقة بين المتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية" بناء على تصنيف الشخص ذكر كان أم أنثى، وأهمية هذه الخطوة تكمن في التحقق من أنه إذا ما وجدت علاقة قوية بين المتغيرين الأساسيين فهل للمتغير الثالث "أثر" على تلك العلاقة أم لا.

في شريط الأدوات العلوي قم باختيار Analyze>Descriptive Statistics>Crosstabs، فتظهر النافذة الخاصة بتكوين الجداول المشتركة (Crosstabs). في تلك النافذة، (شكل (24.4))، قم بنقل المتغير "الفئة العمرية" إلى مربع الصفوف ((Row(s))، ونقل المتغير "مستويات_فيتامين_دال" إلى مربع الأعمدة ((Column(s)) كما نفذناه سابقا، ولتعريف المتغير الثالث، متغير "النوع"، قم بنقل هذا المتغير إلى مربع الطبقات (Layer 1 of 1) في أسفل النافذة كما هو موضح في الشكل.



شكل 24.4: تعريف متغير تصنيف إضافي في النافذة الخاصة بتكوين جداول الاقتران.

وننوه هنا إلى أنه يمكن إضافة أكثر من متغير تصنيف في هذا المربع باستخدام زر التالي (Next) الظاهر في النافذة إذا ما تتطلب الأمر، إلا أنه في مثالنا الحالي لدينا متغير تصنيف إضافي واحد فقط وهو "النوع".

في خيار الإحصاءات (Statistics) قم باختيار اختبار مربع كاي، ومعمل الاقتران ومعاملي فاي وكرامر V ومعاملي كندل تاو B و C ثم اضغط استمرار. وفي نافذة جدول الاقتران، (شكل (24.4))، قم من جديد بالنقر على خيار الخلايا (Cells) وتأكد من عدم اختيار إظهار النسب لعدم الحاجة لها حاليا، ثم اضغط استمرار وفي النافذة الأصلية اختر عرض الأعمدة البيانية المصنفة (Display clustered bar charts)، ثم اضغط موافق.

في نافذة المخرجات، نلاحظ أن جدول الاقتران، (جدول (19.4))، قد تم تقسيمه بحسب الذكور (القسم الأول من الجدول)، والإناث (القسم الثاني)، والكل (القسم الثالث).

ويمكن من هذا الجدول مراقبة التكرارات، (أو النسب إذا ما أردنا)، للذكور على حده أو الإناث. إلا أنه من عيوب هذا التقسيم أن بعض الخلايا ستكون فيها التكرارات مساوية للصفر وخاصة في العينات الصغيرة كما هو الحال في جدولنا الحالي.

وأهم ما يُلاحظ في الجدول (19.4) أن التكرار الأعلى للذكور (10) كان للذكور الذين أعمارهم 51 سنة فأكثر ومستوى فيتامين دال لديهم أقل من 30 نانومول. وأما بالنسبة للإناث، فمعدلات الفيتامين التي تقل عن 30 نانومول كانت متوزعة بصورة متقاربة للفئات العمرية الثلاثة.

جدول 19.4: جدول الاقتران للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية" باستخدام متغير "النوع" كمتغير تصنيف، (ملف البيانات "فيتامين دال").

فئات العمر * مستويات فيتامين دال * النوع Crosstabulation

Count

النوع			أقل من 30	من 30 إلى 75	أكثر من 75	Total
			مستويات فيتامين دال			
ذكر	فئات العمر	أقل من 30	0	0	6	6
		من 30 إلى 50	0	5	5	10
		51 فأكثر	10	3	2	15
	Total		10	8	13	31
أنثى	فئات العمر	أقل من 30	7	2	1	10
		من 30 إلى 50	6	4	0	10
		51 فأكثر	9	0	0	9
	Total		22	6	1	29
Total	فئات العمر	أقل من 30	7	2	7	16
		من 30 إلى 50	6	9	5	20
		51 فأكثر	19	3	2	24
	Total		32	14	14	60

أما الجدول التالي في نافذة المخرجات، (جدول (20.4))، فنلاحظ فيه ما يلي؛

جدول 20.4: نتيجة اختبار مربع كاي للاستقلالية للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية" باستخدام متغير "النوع" كمتغير تصنيف، (ملف البيانات "فيتامين دال").

Chi-Square Tests

النوع		Value	df	Asymptotic Significance (2-sided)
ذكر	Pearson Chi-Square	22.584[b]	4	**.000**
	Likelihood Ratio	27.207	4	.000
	Linear-by-Linear Association	15.888	1	.000
	N of Valid Cases	31		
أنثى	Pearson Chi-Square	6.635[c]	4	**.156**
	Likelihood Ratio	8.300	4	.081
	Linear-by-Linear Association	2.627	1	.105
	N of Valid Cases	29		
Total	Pearson Chi-Square	16.552[a]	4	**.002**
	Likelihood Ratio	16.285	4	.003
	Linear-by-Linear Association	7.956	1	.005
	N of Valid Cases	60		

a. 4 cells (44.4%) have expected count less than 5. The minimum expected count is 3.73.
b. 8 cells (88.9%) have expected count less than 5. The minimum expected count is 1.55.
c. 6 cells (66.7%) have expected count less than 5. The minimum expected count is .31.

من الصف الأول في القسم الأول من الجدول، الخاص بالذكور، نرى أن $\chi_c^2 = 22.584$ وأن القيمة الاحتمالية (P-value = 0) وهذا يدفعنا لرفض الفرضية الصفرية؛ توجد استقلالية بين المتغيرين (للذكور) H_0:، والقول بوجود علاقة ذات أهمية بين معدلات فيتامين دال وأعمار الرجال في العموم. ومن الصف الأول في القسم الثاني من الجدول، الخاص بالإناث، نرى أن $\chi_c^2 = 6.635$ وأن القيمة الاحتمالية (P-value = 0.156) وهذا يدفعنا لقبول الفرضية الصفرية؛ توجد استقلالية بين المتغيرين (للإناث) H_0: ، والقول بعدم وجود علاقة بين معدلات فيتامين دال وأعمار النساء في العموم. وتتأكد هذه النتائج من خلال حساب معاملات الارتباط في جدول (21.4).

جدول 21.4: معاملات الارتباط لجدول الاقتران للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية" باستخدام متغير "النوع" كمتغير تصنيف، (ملف البيانات "فيتامين دال").

Symmetric Measures

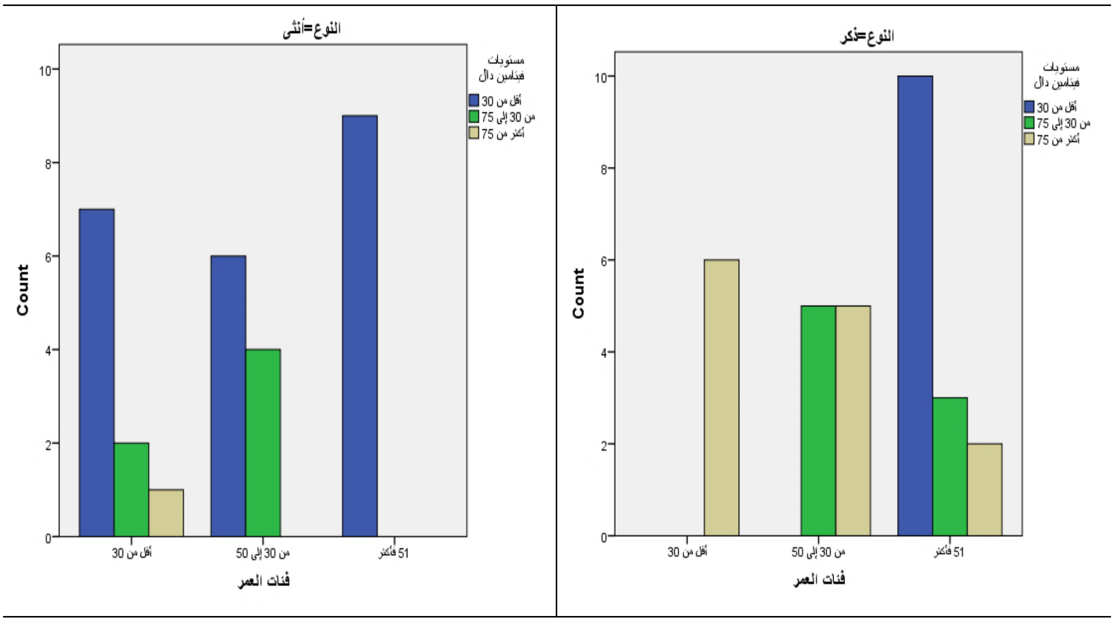
النوع			Value	Asymptotic Standardized Error[a]	Approximate T[b]	Approximate Significance
ذكر	Nominal by Nominal	Phi	.854			**.000**
		Cramer's V	.604			**.000**
		Contingency Coefficient	.649			**.000**
	Ordinal by Ordinal	Kendall's tau-b	-.694	.089	-7.148	**.000**
		Kendall's tau-c	-.665	.093	-7.148	**.000**
	N of Valid Cases		31			
أنثى	Nominal by Nominal	Phi	.478			**.156**
		Cramer's V	.338			**.156**
		Contingency Coefficient	.431			**.156**
	Ordinal by Ordinal	Kendall's tau-b	-.269	.136	-1.797	**.072**
		Kendall's tau-c	-.203	.113	-1.797	**.072**
	N of Valid Cases		29			
Total	Nominal by Nominal	Phi	.525			**.002**
		Cramer's V	.371			**.002**
		Contingency Coefficient	.465			**.002**
	Ordinal by Ordinal	Kendall's tau-b	-.347	.114	-3.020	**.003**
		Kendall's tau-c	-.329	.109	-3.020	**.003**
		N of Valid Cases	60			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

حيث نلاحظ من الجدول، (في القسم الأول)، أن جميع معاملات الارتباط للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية" قيمها مرتفعة عند الذكور، إضافة لكونها ذات معنوية، (كل القيم الاحتمالية P-value=0)، وهذا يدل على وجود علاقة قوية بين المتغيرين عند الرجال.

أما عند النساء، (في القسم الثاني من الجدول (21.4))، فالوضع على النقيض، إذ نلاحظ أن جميع القيم الاحتمالية P-value>0.05، مما يدفعنا للقول بعدم وجود علاقة معنوية بين "مستويات_فيتامين_دال" و"الفئة_العمرية" عند النساء.

وهذه النتائج يمكن ملاحظتها بصورة مرئية من خلال الأعمدة البيانية المصنفة لكل من الذكور والإناث في الشكل (25.4).



شكل 25.4: الأعمدة البيانية المصنفة للمتغيرين "الفئة_العمرية" و"مستويات_فيتامين_دال" للذكور (إلى اليمين)، والإناث (إلى اليسار)، (ملف البيانات "فيتامين دال").

وكخلاصة لما سبق من النتائج الخاصة بدراسة العلاقة بين متغيري "مستويات_فيتامين_دال" و"الفئة_العمرية"، (والتي تشمل جدول الاقتران المتضمن للنسب، الأعمدة البيانية المصنفة، اختبار الاستقلالية، ومعاملات الارتباط لجدول الاقتران بوجود متغير تصنيف النوع)، نستطيع القول بوجود علاقة عكسية قوية بين معدلات فيتامين دال وأعمار الرجال، حيث أن الرجال الذين اجتازوا الخمسين عاما هم أكثر عرضة لنقص فيتامين دال في أجسامهم.

4.4 تحليل التباين (Analysis of Variance (ANOVA))

يُعد مفهوم تحليل التباين من المفاهيم الهامة والأساسية في نظرية الاستدلال الإحصائي، حيث أنه يعتمد على مقارنة تساوي متوسطات عدة مجتمعات، (ثلاثة فأكثر)، من خلال تحليل الاختلاف بين تلك المتوسطات، وهو ما يُعرف بتحليل التباين في اتجاه واحد (One-way ANOVA)، أو تحليل الاختلاف بين المتوسطات وداخل المجتمعات (أو المجموعات) معا، وهو ما يُعرف بتحليل التباين في اتجاهين (Two-way ANOVA). وسنبدأ بشرح كيفية تنفيذ التحليل الأول؛

1.4.4 تحليل التباين في اتجاه واحد (One-way ANOVA)

سنستخدم ملف البيانات "معدلات الطلبة" لتوضيح كيفية تنفيذ تحليل التباين في اتجاه واحد في SPSS، حيث يمثل "معدل_الطالب" المتغير الرئيسي أو التابع في التحليل ويمثل "القسم_العلمي" متغير تقسيم المجموعات والتي تمثل الأقسام العلمية الثلاثة؛ الرياضيات (MA)، الإحصاء (ST)، والفيزياء (PH). وسيكون الهدف هو

التحقق من تجانس/اختلاف المستوى الدراسي للطلبة في الأقسام الثلاثة، أي اختبار الفرضية الصفرية:

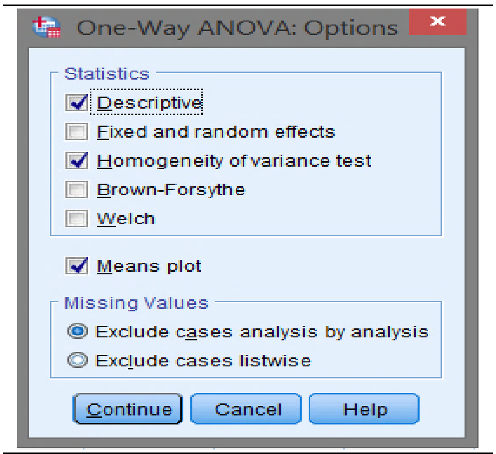
$$H_0: \mu_{MA} = \mu_{ST} = \mu_{PH}$$

في شريط الأوامر العلوي، قم باختيار Analyze>Compare Means>One-Way ANOVA، فتظهر نافذة تحليل التباين في اتجاه واحد (One-way ANOVA) كما يوضح الشكل (26.4). قم في تلك النافذة بنقل المتغير "معدل_الطالب" إلى المربع الخاص بالمتغير التابع (Dependent List) ونقل المتغير "القسم_العلمي" إلى المربع الخاص بمتغير التصنيف أو المجموعات (Factor).



شكل 26.4: نافذة تحليل التباين في اتجاه واحد لملف البيانات "معدلات الطلبة".

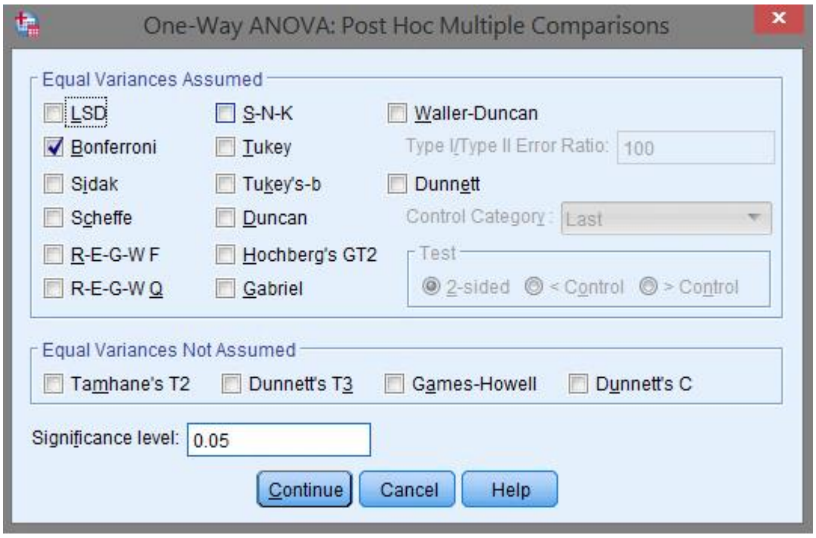
اضغط على أمر الخيارات (Options) فتظهر النافذة الفرعية التالية، (شكل (27.4))، قم فيها باختيار عرض المقاييس الإحصائية الوصفية (Descriptive) واختيار تساوي التباينات (Homogeneity of variance test) والتمثيل البياني للمتوسطات (Means plot)، وهذه الخيارات في الواقع قد تم التطرق إليها في مواضع سابقة في الكتاب إلا أننا نقوم باختيارها هنا لتكون جزء مكملا لنتائج تحليل التباين. بعد الانتهاء من اختيار ما سبق اضغط استمرار للعودة للنافذة الأصلية للتحليل.



شكل 27.4: نافذة الخيارات ضمن نافذة تحليل التباين في اتجاه واحد.

حيث أننا نقوم حاليا باختبار تساوي ثلاثة متوسطات، فمن المتوقع أن تكون إما كلها متساوية، (وهذا ما ينتج عن قبول الفرضية الصفرية لاختبار تحليل التباين)، أو أن نرفض تلك الفرضية الصفرية والذي سينتج عنه وجود متوسطان متساويان على الأقل، بمعنى أننا في مثالنا الحالي سنكون بحاجة لاختبار تساوي المتوسطات في ثلاثة حالات هي؛ $H_0: \mu_{MA} = \mu_{ST}$ ، $H_0: \mu_{MA} = \mu_{PH}$، و $H_0: \mu_{ST} = \mu_{PH}$، بمعنى اختبار تساوي المستوى الدراسي للطلبة في كل قسمين على حده، وهو ما يتم باستخدام اختبار تساوي الأوساط في مجتمعين كما شرحنا سابقا في هذا الفصل.

لذلك، وكخطوة استباقية، سنقوم باختيار تنفيذ هذه الاختبارات "الفرعية" للحصول على معلومات إضافية تساعدنا في الحكم النهائي على مدى تجانس المستوى الدراسي للطلبة في الأقسام العلمية. ولعمل ذلك قم باختيار خيار التحليل البَعدي (Post Hoc)، والذي سيُظهر النافذة الفرعية التالية، (شكل (28.4))؛



شكل 28.4: نافذة التحليل البَعدي في تحليل التباين في اتجاه واحد.

قم في تلك النافذة باختيار **اختبارات المقارنة المتعددة لبونفيروني** (Bonferroni)، ويمكن التحكم بمستوى المعنوية المطلوب من مربع (Significance level) في أسفل النافذة، وسنبقي هنا على مستوى المعنوية الافتراضي وهو 0.05. قم بعدها بالضغط على استمرار للعودة للنافذة الأصلية، وفي تلك النافذة اضغط موافق.

ستظهر بعد ذلك النتائج المطلوبة في نافذة المخرجات، والجدول الأول في هذه النتائج، (جدول (22.4))، يضم القيم المحسوبة للمعدلات في كل قسم من الأقسام الثلاثة، وهذه القيم تشمل الوسط الحسابي، الانحراف المعياري، الخطأ المعياري لتقدير الوسط الحسابي، فترة الثقة لتقدير الوسط الحسابي للمجتمع، والقيمة الصغرى والكبرى للمعدلات. وهذه المقاييس قد تم التعليق عليها في الفصل الثالث لنفس البيانات.

جدول 22.4: بعض المقاييس الإحصائية لمعدلات الطلبة في الأقسام العلمية لملف البيانات "معدلات الطلبة".

Descriptives

معدل الطالب

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
MA	30	2.0027	.67117	.12254	1.7520	2.2533	1.26	3.67
ST	30	1.6463	.25022	.04568	1.5529	1.7398	1.15	2.18
PH	30	1.6857	.21144	.03860	1.6067	1.7646	1.26	2.31
Total	90	1.7782	.45550	.04801	1.6828	1.8736	1.15	3.67

والجدول الثاني في نافذة المخرجات، (جدول (23.4))، يمثل نتيجة اختبار تجانس تباينات معدلات الطلبة في الأقسام العلمية الثلاثة باستخدام اختبار ليفين، ومنه نستنتج وجود اختلاف بين تباينات معدلات الطلبة، (P-value = 0).

جدول 23.4: اختبار تجانس التباينات لليفن لمعدلات الطلبة في الأقسام في ملف البيانات "معدلات الطلبة".

Test of Homogeneity of Variances

معدل الطالب

Levene Statistic	df1	df2	Sig.
14.098	2	87	**.000**

أما الجدول الثالث، (جدول (24.4))، فيمثل النتيجة الرئيسية في هذا التحليل الإحصائي وهو جدول تحليل التباين في اتجاه واحد.

جدول 24.4: جدول تحليل التباين في اتجاه واحد لمعدلات الطلبة في الأقسام الثلاثة في ملف البيانات "معدلات الطلبة".

ANOVA

معدل الطالب

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.290	2	1.145	6.159	**.003**
Within Groups	16.176	87	.186		
Total	18.466	89			

ومن هذا الجدول يمكننا الجزم بوجوب رفض الفرضية الصفرية القائلة بتساوي المتوسطات الثلاثة، (حيث أن القيمة الاحتمالية P-value =0.003)، وهذا يدل على وجود متوسطان مختلفان على الأقل، (أي وجود قسمان على الأقل يختلف فيهما مستوى الطلبة الدراسي)، ضمن المتوسطات الثلاثة. وهذا يدفعنا لإجراء اختبارات المقارنة المتعددة للمتوسطات في الجدول التالي، جدول (25.4)؛

جدول 25.4: جدول اختبارات المقارنة المتعددة للمتوسطات في أقسام الرياضيات، الإحصاء، والفيزياء في ملف البيانات "معدلات الطلبة".

Multiple Comparisons

Dependent Variable: معدل الطالب
Bonferroni

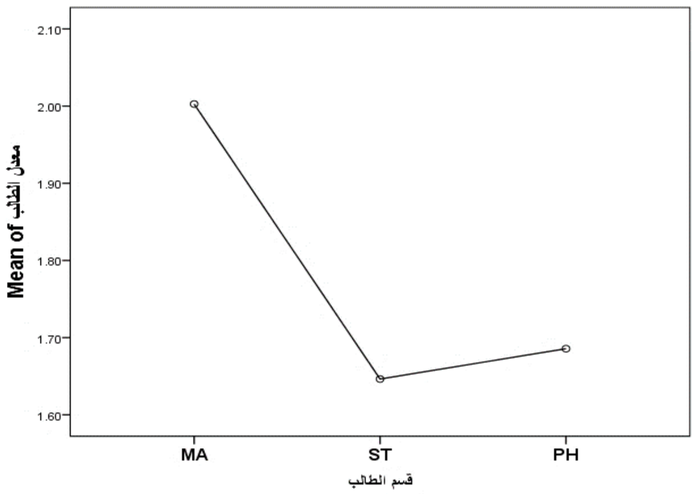
قسم الطالب (I)		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
MA	ST	.35633*	.11133	**.006**	.0846	.6281
	PH	.31700*	.11133	**.017**	.0452	.5888
ST	MA	-.35633-*	.11133	**.006**	-.6281	-.0846
	PH	-.03933	.11133	**1.000**	-.3111	.2324
PH	MA	-.31700-*	.11133	**.017**	-.5888	-.0452
	ST	.03933	.11133	**1.000**	-.2324	.3111

*. The mean difference is significant at the 0.05 level.

ومن هذا الجدول يتبين لنا ما يلي:

1. وجود اختلاف بين معدلات الطلبة في قسمي الرياضيات والإحصاء، حيث أن القيمة الاحتمالية لاختبار $H_0: \mu_{MA} = \mu_{ST}$ هي 0.006 مما يدفعنا لرفض هذه الفرضية الصفرية.
2. وجود اختلاف بين معدلات الطلبة في قسمي الرياضيات والفيزياء أيضا، حيث أن القيمة الاحتمالية لاختبار $H_0: \mu_{MA} = \mu_{PH}$ هي 0.017 مما يجعلنا نرفض الفرضية الصفرية.
3. وجود تكافؤ بين معدلات الطلبة في قسمي الإحصاء والفيزياء، حيث أن القيمة الاحتمالية لاختبار الفرضية $H_0: \mu_{ST} = \mu_{PH}$ هي 1.00 وبالتالي نقبل الفرضية الصفرية.

والنتيجة الأخيرة في نافذة المخرجات، شكل (29.4)، تُظهر التمثيل البياني لمتوسطات المعدلات في الأقسام العلمية الثلاثة، ونلاحظ فيها بوضوح مدى اختلاف (ارتفاع) متوسط معدلات الطلبة في قسم الرياضيات عن قسمي الإحصاء والفيزياء.



شكل 29.4: التمثيل البياني لمتوسطات الأقسام الثلاثة.

2.4.4 تحليل التباين في اتجاهين (Two-way ANOVA)

في تحليل التباين في اتجاه واحد، رأينا كيف أن الاهتمام كان متركزا على دراسة الاختلاف بين عدة مجموعات، (والتي تكون عادة ممثلة بأعمدة البيانات)، أما في تحليل التباين في اتجاهين، فإنه إضافة لاختبار الأعمدة واختبار الصفوف، (كاتجاه واحد أيضا)، فإننا نقوم بدراسة تداخل الأعمدة والصفوف معا، أي كأننا ندرس العلاقة بين متغيرين. ولتوضيح كيفية تنفيذ تحليل التباين في اتجاهين لنأخذ المثال التالي؛

البيانات في الجدول (26.4) تمثل أطوال عينة مكونة من 20 طفلا بالسنتيمتر تم تقسيمهم إلى خمسة مجموعات (أعمدة الجدول)، كل مجموعة تم إعطاؤها نوع محدد من الحليب (Milk1, Milk2, …, Milk5)، وبداخل كل مجموعة من هذه المجموعات الخمسة تم اعتماد أربعة أنواع من فيتامين دال (Vit.D1, …, Vit.D4) بواقع نوع واحد من الفيتامين لكل طفل، (وهو ما تمثله صفوف الجدول)، وذلك خلال عدة سنوات، بمعنى أن الطفل الأول في العينة، (القيمة الأولى في أعلى يسار الجدول)، والذي طوله 95 سم، قد تم إعطاؤه نوع الحليب (Milk1) ونوع الفيتامين (Vit.D1)، وهكذا.

جدول 26.4: أطوال عينة من الأطفال بالسنتيمتر بحسب نوع الحليب ونوع الفيتامين المستخدَم.

		نوع حليب الأطفال				
		Milk1	Milk2	Milk3	Milk4	Milk5
نوع فيتامين دال	Vit.D1	95	101	93	96	100
	Vit.D2	97	100	98	97	94
	Vit.D3	120	116	125	118	122
	Vit.D4	90	88	93	89	87

سنقوم أولا بإدخال هذه البيانات في ملف جديد في SPSS باسم "أطوال الأطفال"، بحيث يتم تعريف ثلاثة متغيرات كالتالي:

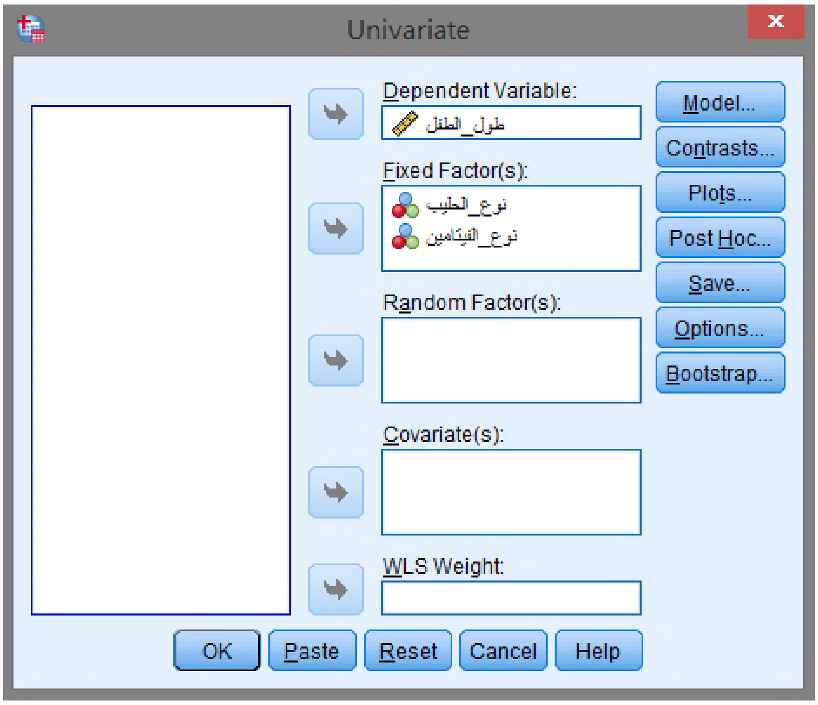
- المتغير الأول، باسم "طول_الطفل"، ويمثل أطوال جميع الأطفال في العينة، حيث يتم إدخال قيم العامود الأول (Milk1) أولا ثم قيم العامود الثاني (Milk2) تحتها، وهكذا حتى العامود الخامس (Milk5).
- المتغير الثاني، باسم "نوع_الحليب"، ويمثل نوع حليب الأطفال المعطى لكل طفل، حيث سيتم إعطاء القيمة 1 لأول أربعة أطفال، والقيمة 2 لثاني أربعة أطفال، وهكذا حتى القيم 5 التي ستعطى لآخر أربعة أطفال. مع ملاحظة إعطاء القيم من 1 إلى 5 التعريف (Milk1) إلى (Milk5) على الترتيب في خانة القيم (Values) في نافذة المتغيرات.
- المتغير الثالث، باسم "نوع_الفيتامين"، ويمثل نوع الفيتامين المعطى لكل طفل، حيث سيتم إعطاء القيم 1، 2، 3، 4 لكل أربعة أطفال بدء من الأعلى، إلى الأسفل. مع ملاحظة إعطاء القيم من 1 إلى 4 التعريف (Vit.D1) إلى (Vit.D4) على الترتيب في خانة القيم (Values) في نافذة المتغيرات.

عند الانتهاء سيأخذ ملف البيانات الشكل التالي، (شكل (30.4))؛

	طول_الطفل	نوع_الحليب	نوع_الفيتامين
1	95	Milk1	Vit.D1
2	97	Milk1	Vit.D2
3	120	Milk1	Vit.D3
4	90	Milk1	Vit.D4
5	101	Milk2	Vit.D1
6	100	Milk2	Vit.D2
7	116	Milk2	Vit.D3
8	88	Milk2	Vit.D4
9	93	Milk3	Vit.D1
10	98	Milk3	Vit.D2
11	125	Milk3	Vit.D3

شكل 30.4: ملف البيانات "أطوال_الأطفال" في SPSS.

ولتنفيذ تحليل التباين في اتجاهين، اختر Analyze>General Linear Model>Univariate في شريط الأدوات العلوي فتظهر النافذة التالية، (شكل (31.4))؛



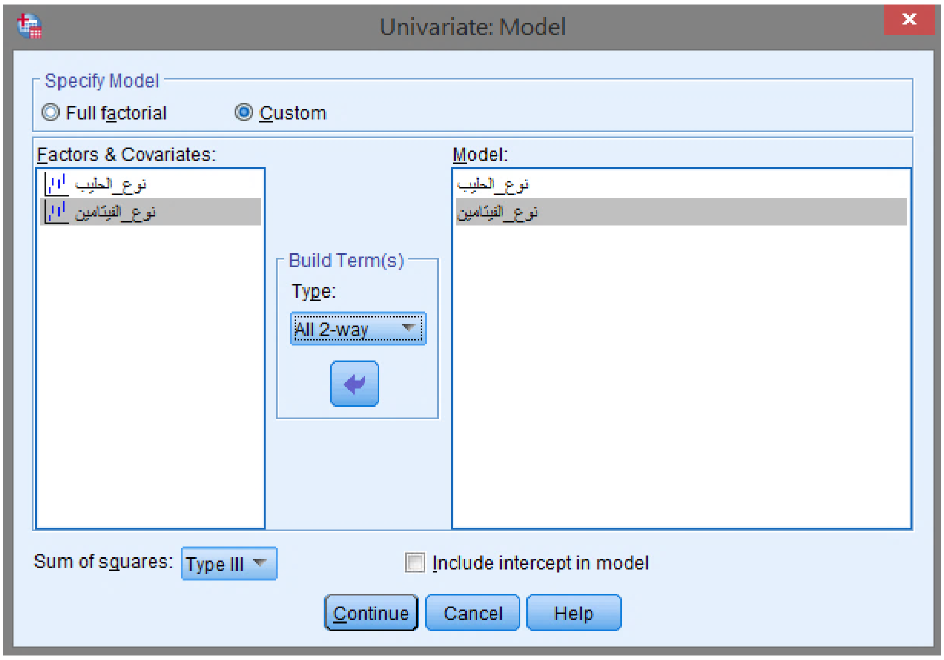
شكل 31.4: نافذة تحليل التباين في اتجاهين لملف البيانات "أطوال_الأطفال".

في هذه النافذة، قم باختيار المتغير "طول_الطفل" كمتغير تابع (Dependent Variable)، والمتغيران "نوع_الحليب" و"نوع_الفيتامين" كمتغيرات تقسيم (Fixed Factor(s)) كما هو موضح في الشكل السابق.

الآن اضغط على خيار النموذج (Model) فتظهر نافذة فرعية، (كما يظهر في الشكل (32.4))، وسنقوم بتنفيذ أربعة أمور في تلك النافذة، كما يظهر فيها:

1. تغير خيار تحديد النموذج (Specify Model) إلى مخصص (Custom).
2. نقل كل من المتغيرين "نوع_الحليب" و"نوع_الفيتامين" إلى مربع النموذج (Model) إلى اليمين.
3. اختيار النوع اتجاهين (All 2-way) من النافذة التي هي في المنتصف (Build Term(s)).
4. اختيار عدم إظهار ثابت النموذج (Include intercept in model) في أسفل النافذة.

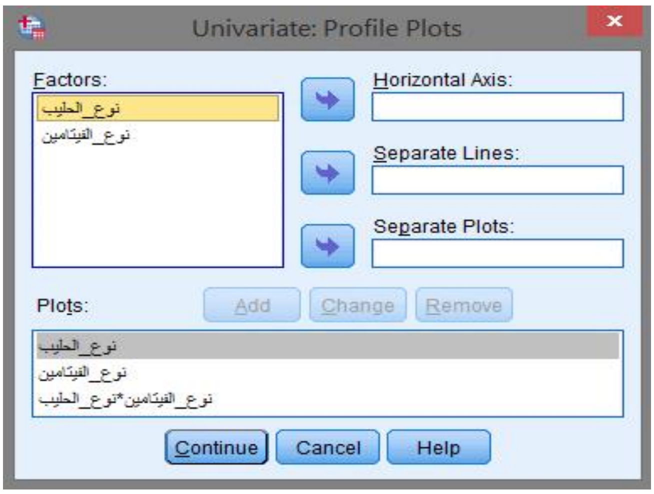
وبعد تنفيذ الخطوات السابقة، اضغط استمرار للعودة للنافذة الأصلية.



شكل 32.4: نافذة اختيار نوع النموذج في تحليل التباين في اتجاهين لملف البيانات "أطوال_الأطفال".

في النافذة الأصلية، (شكل (31.4))، قم بالنقر على خيار الرسم (Plots) لاختيار عرض التغير في المتوسطات لأطوال الأطفال مرة مع نوع الحليب ومرة مع نوع الفيتامين ومرة أخرى مع الاثنين في آن واحد. لتنفيذ ذلك، قم بالخطوات الثلاثة التالية، (في الشكل (33.4)):

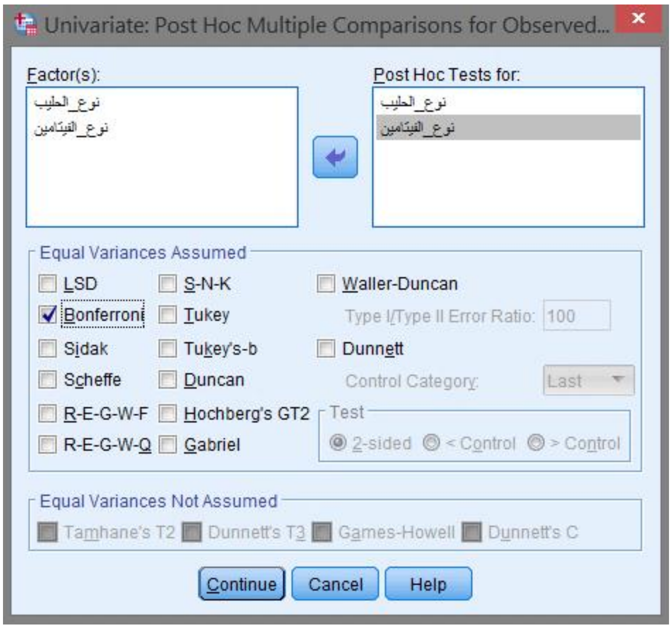
1. انقل المتغير "نوع_الحليب" إلى المحور الأفقي (Horizontal Axis) ثم اضغط إضافة (Add) لاختيار عرض رسم هذا المتغير.



شكل 33.4: نافذة اختيار عرض الرسم في تحليل التباين في اتجاهين لملف البيانات "أطوال_الأطفال".

2. بالمثل انقل المتغير "نوع_الفيتامين" إلى المحور الأفقي (Horizontal Axis) ثم اضغط إضافة (Add) لاختيار عرض رسم هذا المتغير أيضا.
3. الآن انقل المتغير "نوع_الحليب" من جديد في خانة المحور الأفقي (Horizontal Axis) والمتغير "نوع_الفيتامين" إلى خانة الخطوط المنفصلة (Separate Lines) ثم اضغط إضافة (Add).

بعد ذلك اضغط استمرار للعودة للنافذة الأصلية، (شكل (31.4))، وفي تلك النافذة اضغط من جديد خيار التحليل البَعدي (Post Hoc)، والذي سيُظهر نافذة فرعية جديدة، (شكل (34.4))، قم بعدها باختيار إجراء مقارنات بين أوساط مستويات المتغيرين "نوع_الحليب" و"نوع_الفيتامين" عن طريق نقلهما إلى مربع اختبارات التحليل البعدي (Post Hoc Tests for)، ثم اختيار إجراء اختبارات المقارنة المتعددة لبونفيروني (Bonferroni)، والضغط على استمرار. وفي النافذة الأصلية اضغط موافق.



شكل 34.4: نافذة اختيار المقارنات المتعددة للأوساط في تحليل التباين في اتجاهين لملف البيانات "أطوال_الأطفال".

ستظهر النتائج المطلوبة بعد ذلك في نافذة المخرجات. الجدول الأول يُظهر تقسيم المستويات في جدول البيانات وعدد القيم في كل مستوى.

الجدول الثاني، (جدول (27.4))، هو الجدول الأساسي في مجموعة النتائج ويمثل جدول تحليل التباين في اتجاهين.

جدول 27.4: نتائج جدول تحليل التباين في اتجاهين لملف البيانات "أطوال_الأطفال".

Tests of Between-Subjects Effects					
Dependent Variable:	طول_الطفل				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	206497.900[a]	8	25812.238	2516.221	.000
نوع_الحليب	11.700	4	2.925	.285	**.882**
نوع_الفيتامين	2668.150	3	889.383	86.699	**.000**
Error	123.100	12	10.258		
Total	206621.000	20			
a. R Squared = .999 (Adjusted R Squared = .999)					

وأهم ما نستنتج منه:

1. عدم وجود اختلاف بين أطوال الأطفال نتيجة لاختلاف أنواع الحليب، أي أن تأثير أنواع الحليب الخمسة متساوي على نمو الأطفال، بمعنى قبول الفرضية الصفرية:

$$H_0: \mu_{Milk1} = \mu_{Milk2} = \mu_{Milk3} = \mu_{Milk4} = \mu_{Milk5}$$

حيث أن القيمة الاحتمالية للاختبار P-value = 0.882.

2. وجود اختلاف ذو معنوية بين أطوال الأطفال نتيجة لاختلاف أنواع الفيتامين المستخدم، أي أن تأثير أنواع الفيتامين الأربعة غير متساوي على نمو الأطفال، بمعنى رفض الفرضية الصفرية:

$$H_0: \mu_{Vit.D1} = \mu_{Vit.D2} = \mu_{Vit.D3} = \mu_{Vit.D4}$$

حيث أن القيمة الاحتمالية للاختبار P-value = 0.

والنتيجة السابقة تدفعنا للتحقق من أنواع الفيتامين التي المختلفة فيما بينها وذلك عن طريق ملاحظة الجدول الخاص بالمقارنات المتعددة للمتغير "نوع_الفيتامين" في الجدول (28.4)؛

جدول 28.4: المقارنات المتعددة بين متوسطات أنواع الفيتامين في تحليل التباين في اتجاهين لملف البيانات "أطوال_الأطفال".

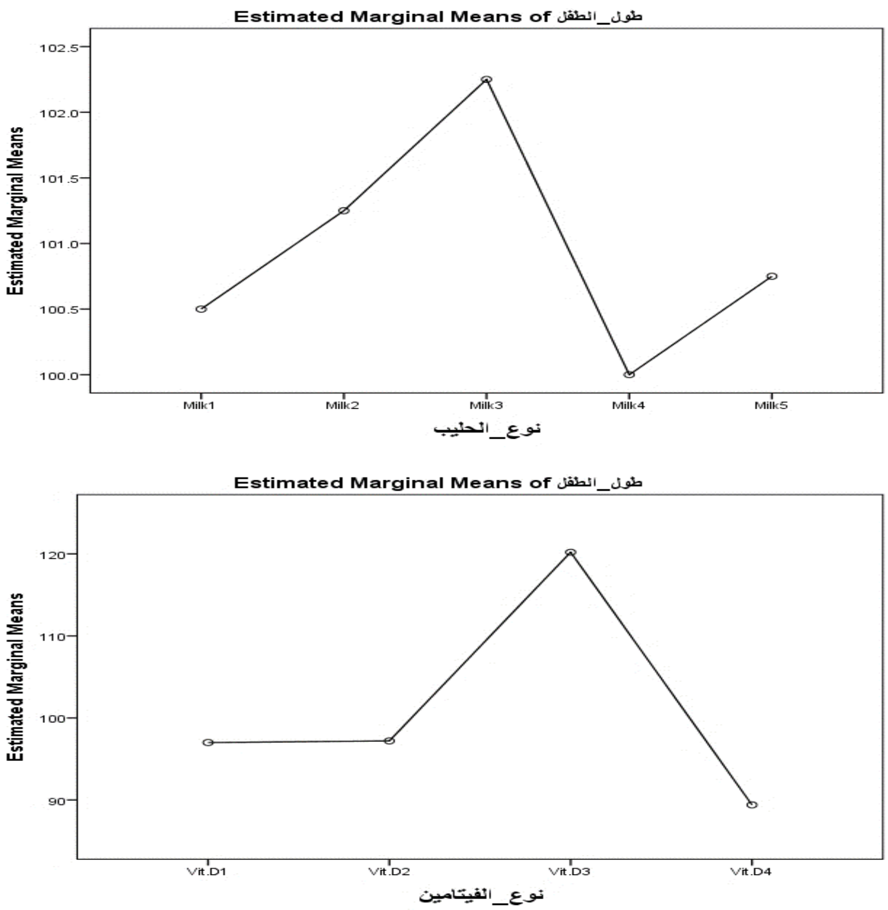
Multiple Comparisons						
Dependent Variable:	طول_الطفل					
Bonferroni						
(I) نوع_الفيتامين		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Vit.D1	Vit.D2	-.20	2.026	**1.000**	-6.59	6.19
	Vit.D3	-23.20-*	2.026	**.000**	-29.59	-16.81
	Vit.D4	7.60*	2.026	**.017**	1.21	13.99
Vit.D2	Vit.D1	.20	2.026	**1.000**	-6.19	6.59
	Vit.D3	-23.00-*	2.026	**.000**	-29.39	-16.61
	Vit.D4	7.80*	2.026	**.014**	1.41	14.19
Vit.D3	Vit.D1	23.20*	2.026	**.000**	16.81	29.59
	Vit.D2	23.00*	2.026	**.000**	16.61	29.39
	Vit.D4	30.80*	2.026	**.000**	24.41	37.19
Vit.D4	Vit.D1	-7.60-*	2.026	**.017**	-13.99	-1.21
	Vit.D2	-7.80-*	2.026	**.014**	-14.19	-1.41
	Vit.D3	-30.80-*	2.026	**.000**	-37.19	-24.41

Based on observed means.
The error term is Mean Square (Error) = 10.258.
*. The mean difference is significant at the 0.05 level.

من هذا الجدول، نستطيع ملاحظة التالي:

1. وجود اختلاف معنوي بين نوع الفيتامين الأول (Vit.D1) وكل من نوعي الفيتامين الثالث والرابع (Vit.D3 وVit.D4)، حيث أن القيمة الاحتمالية الخاصة باختبار تساوي المتوسطات هي أقل من 0.05.
2. عدم وجود اختلاف معنوي بين نوع الفيتامين الأول (Vit.D1) ونوع الفيتامين الثاني (Vit.D2)، حيث أن القيمة الاحتمالية الخاصة باختبار تساوي المتوسطات هي أكبر من 0.05.
3. وجود اختلاف معنوي بين نوع الفيتامين الثاني (Vit.D2) وكل من نوعي الفيتامين الثالث والرابع (Vit.D3 وVit.D4)، حيث أن القيمة الاحتمالية الخاصة باختبار تساوي المتوسطات هي أقل من 0.05.
4. وجود اختلاف معنوي بين نوع الفيتامين الثالث (Vit.D3) وكل أنواع الفيتامين الأخرى، حيث أن القيمة الاحتمالية الخاصة باختبار تساوي المتوسطات هي أقل من 0.05.

وهذه النتائج يمكن ملاحظتها أيضا بصورة مرئية من خلال التمثيل البياني، شكل (35.4)، الخاص بالتغير في متوسطات أنواع الحليب ومتوسطات أنواع الفيتامين.



شكل 35.4: التمثيل البياني للأوساط في تحليل التباين في اتجاهين للمتغيرات في ملف البيانات "أطوال_الأطفال".

الفصل الخامس

تحليل الارتباط والانحدار الخطي
(Linear Correlation and Regression Analysis)

إن من أهم المفاهيم التي يرتكز عليها علم الإحصاء هو دراسة العلاقات ضمن المتغيرات في البيانات، وتحليلها، وتكوين النماذج التي تفسر العلاقات السببية بينها. وفي هذا الفصل، سنستعرض كيفية تنفيذ تحليل الارتباط والانحدار الخطي باستخدام برنامج SPSS.

1.5 تحليل الارتباط الخطي (Linear Correlation Analysis)

يستخدم تحليل الارتباط لدراسة العلاقات بين المتغيرات، وذلك من خلال حساب معامل (مقياس كمي) يسمى **بمعامل الارتباط** (Correlation Coefficient) والذي بدوره يعكس من خلال قيمته نوع وقوة العلاقة الموجودة بين المتغيرات. ويمكن تصنيف دراسة الارتباط من حيث طبيعة العلاقة بين المتغيرات إلى **الارتباط الخطي وغير الخطي** (Linear and Non Linear Correlation).

فالارتباط الخطي هو الذي يهتم بدراسة العلاقات الخطية بين متغيرات الدراسة والتي يمكن تمثيلها بمعادلات رياضية من الدرجة الأولى؛ (مثلا: $y = 3x - 1$)، وهذا النوع من الارتباط هو الذي سيتم تناوله في هذا الفصل من الكتاب. أما الارتباط غير الخطي فهو يدرس العلاقات التي تمثلها معادلات غير خطية؛ (مثلا: $y = 3x^2 - 1$ أو $y = 5e^x$).

وكذلك يمكن تصنيف دراسة الارتباط من حيث طبيعة معامل الارتباط وعدد المتغيرات إلى **الارتباط البسيط، الجزئي، والمتعدد** (Simple, Partial, and Multiple Correlation)؛ فالارتباط البسيط هو الذي يدرس العلاقة بين متغيرين فقط، والجزئي يدرس العلاقة بين متغيرين بوجود تأثير لمتغير أو أكثر على هذين المتغيرين، أما الارتباط المتعدد فهو يقيس العلاقة بين مجموعة من المتغيرات في آن واحد.

ويتم عادة مراقبة أو دراسة الارتباط بين المتغيرات بطريقتين؛ طريقة الرسم وهي تمثيل قيم المتغيرات بيانيا وهو ما يعرف **بشكل الانتشار** (Scatter Plot). والطريقة الثانية هي الطريقة الجبرية وذلك بحساب قيمة معامل الارتباط من خلال الصيغة الحسابية الخاصة بالمعامل.

1.1.5 دراسة الارتباط باستخدام شكل الانتشار (Analyzing Correlation by Scatterplot)

إن دراسة الارتباط عن طريق التمثيل البياني لمتغيرين باستخدام شكل الانتشار يتم عن طريق رسم النقاط على محوري الرسم السيني (X) والصادي (Y) في بُعدين (2D)، ويعد هذا انعكاسا لقيمة معامل الارتباط الخطي في الواقع. ولتوضيح كيفية تنفيذ ذلك في SPSS لنأخذ المثال التالي؛

قم في برنامج SPSS بفتح ملف البيانات "بيانات الطلبة" ونذكر القارئ هنا بأن هذه البيانات تضم ثمانية متغيرات لعينة مكونة من 35 طالبا في إحدى الكليات، وهذه المتغيرات هي تقديرات الطلبة في ثلاثة مقررات دراسية؛ 1، 2، و3، عمر الطالب بالسنوات، جنس الطالب، ترتيب الفصل الدراسي للطالب، عدد أفراد أسرة الطالب، وعدد الحجرات في منزل الطالب.

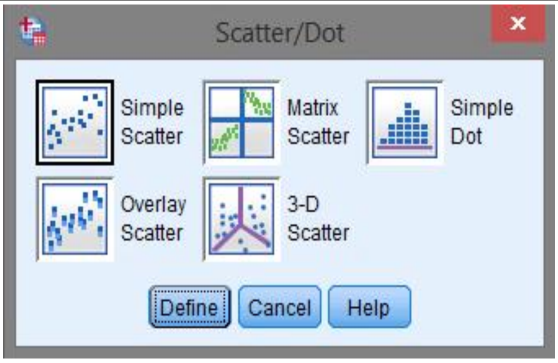
سنقوم أولا بإضافة متغير جديد لهذه البيانات بهدف تعزيز مفهوم العلاقة بين المتغيرات. هذا المتغير، (جدول (1.5))، هو معدل عدد ساعات الدراسة اليومية للطالب خارج مواعيد المحاضرات، أي تقدير الطالب لعدد الساعات التي يستغرقها في الدراسة والمراجعة خلال اسبوع مقسوما على سبعة أيام.

جدول 1.5: متغير معدل عدد ساعات الدراسة اليومية للطالب.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.5	**1**	**2**	**2.5**	**0.5**	**3**	**3.5**	**5**	**4.5**	**3.5**	**1**	**2**	**2.5**	**3.5**	**4**	**1**	**0.5**	**1**
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	
4.5	**3**	**3.5**	**4.5**	**2.5**	**4**	**1.5**	**3.5**	**5**	**2**	**2.5**	**2.5**	**2.5**	**1.5**	**5**	**2**	**5.5**	

بعد إضافة هذا المتغير، والذي سيأخذ الاسم "ساعات_الدراسة"، لملف البيانات "بيانات الطلبة" سنقوم بحفظ الملف باسم "بيانات الطلبة2" في برنامج SPSS.

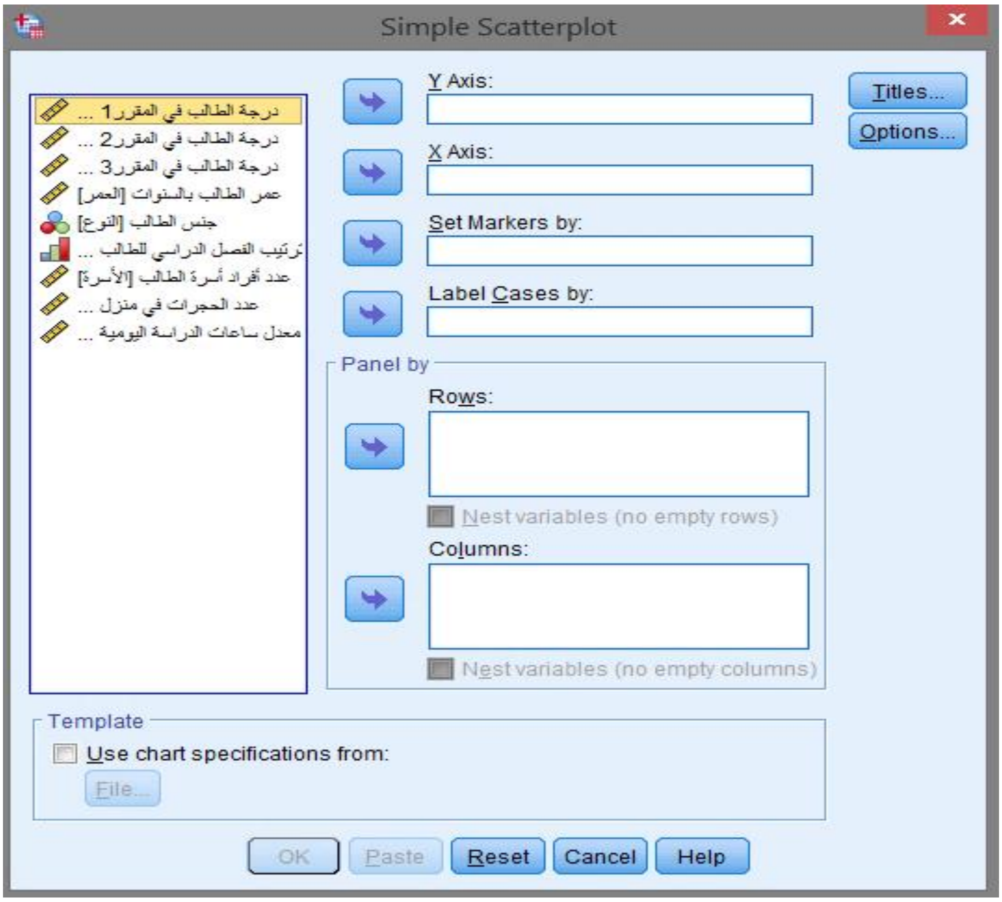
لتنفيذ شكل الانتشار، قم من شريط الأوامر العلوي باختيار Graphs>Legacy Dialogs>Scatter/Dot فتظهر النافذة التالية، (شكل (1.5))؛ في تلك النافذة، سنبقي على الخيار الافتراضي، وهو شكل الانتشار البسيط (Simple Scatter)، على أن نتطرق للخيارات الأخرى لشكل الانتشار لاحقا.



شكل 1.5: نافذة اختيار نوع شكل الانتشار.

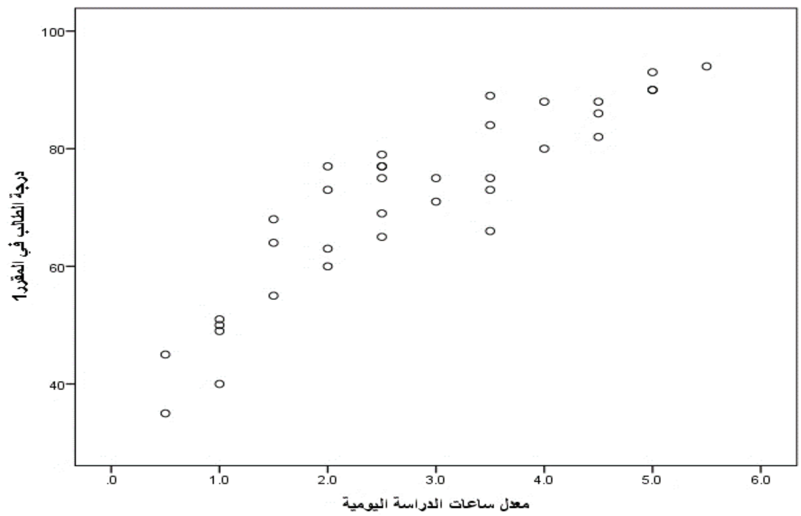
نضغط بعد ذلك على تعريف (Define) فتظهر نافذة جديدة لاختيار المتغيرات في شكل الانتشار البسيط كما في الشكل (2.5).

في تلك النافذة، سنقوم باختيار المتغير "مقرر1" ونقله لمربع المحور الصادي (Y-Axis) واختيار المتغير "ساعات_الدراسة" ونقله لمربع المحور السيني (X-Axis)، أي أننا نريد التعرف على شكل العلاقة بين درجات الطلبة في المقرر 1 ومعدل ساعات دراستهم اليومية.



شكل 2.5: نافذة اختيار متغيرات شكل الانتشار البسيط.

بعد اختيار المتغيرات والضغط على موافق سيظهر شكل الانتشار، (شكل (3.5))، في نافذة المخرجات.

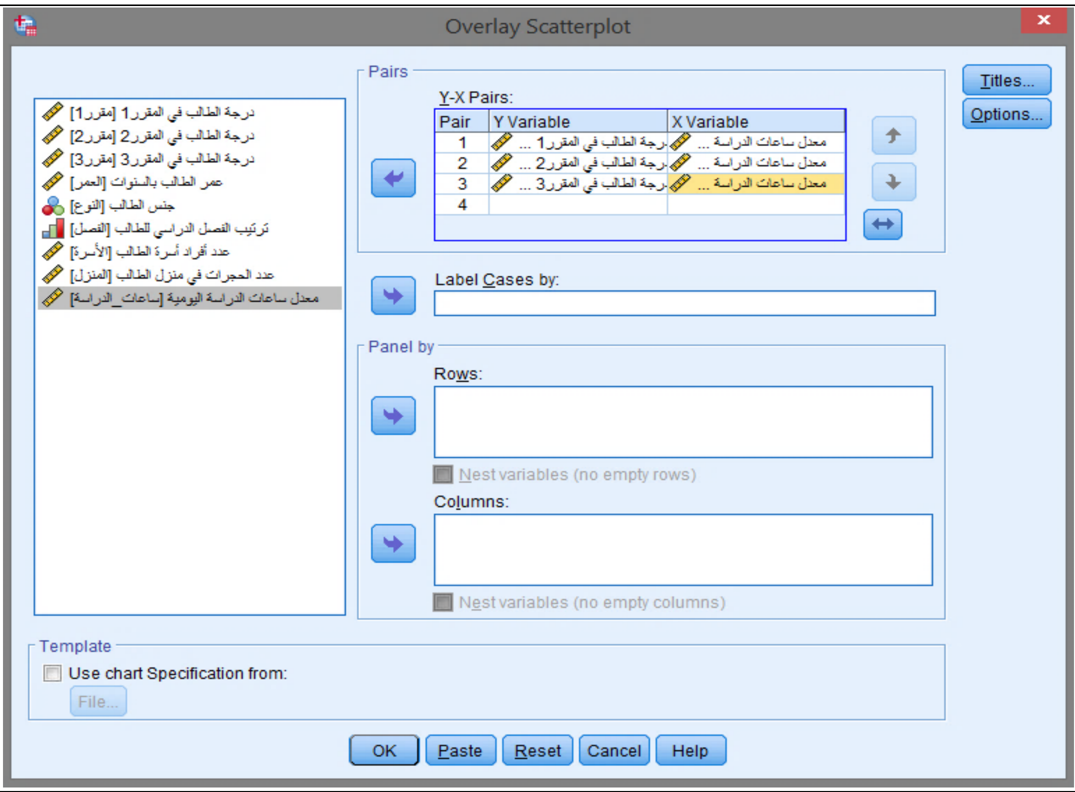


شكل 3.5: شكل الانتشار البسيط للمتغيرين "مقرر1" و"ساعات_الدراسة".

من شكل (3.5)، يمكننا بوضوح ملاحظة وجود علاقة خطية طردية قوية جدا بين المتغيرين، وهذا يكون منطقيا لأنه بزيادة ساعات الدراسة اليومية فإن الدرجات ستزيد في المقرر.

ويمكننا أيضا استخدام شكل الانتشار لمراقبة علاقة متغير بأكثر من متغير على نفس الرسم، فمثلا إذا ما أردنا التعرف على شكل علاقة معدل ساعات الدراسة بدرجات الطلبة في المقررات الثلاثة، يمكننا عمل الآتي:

في شريط الأوامر العلوي باختيار Graphs>Legacy Dialogs>Scatter/Dot وفي نافذة اختيار شكل الانتشار، (شكل (1.5))، قم باختيار شكل الانتشار الفوقي (Overlay Scatter) ثم اضغط تعريف فتظهر نافذة جديدة كما في الشكل (4.5)؛

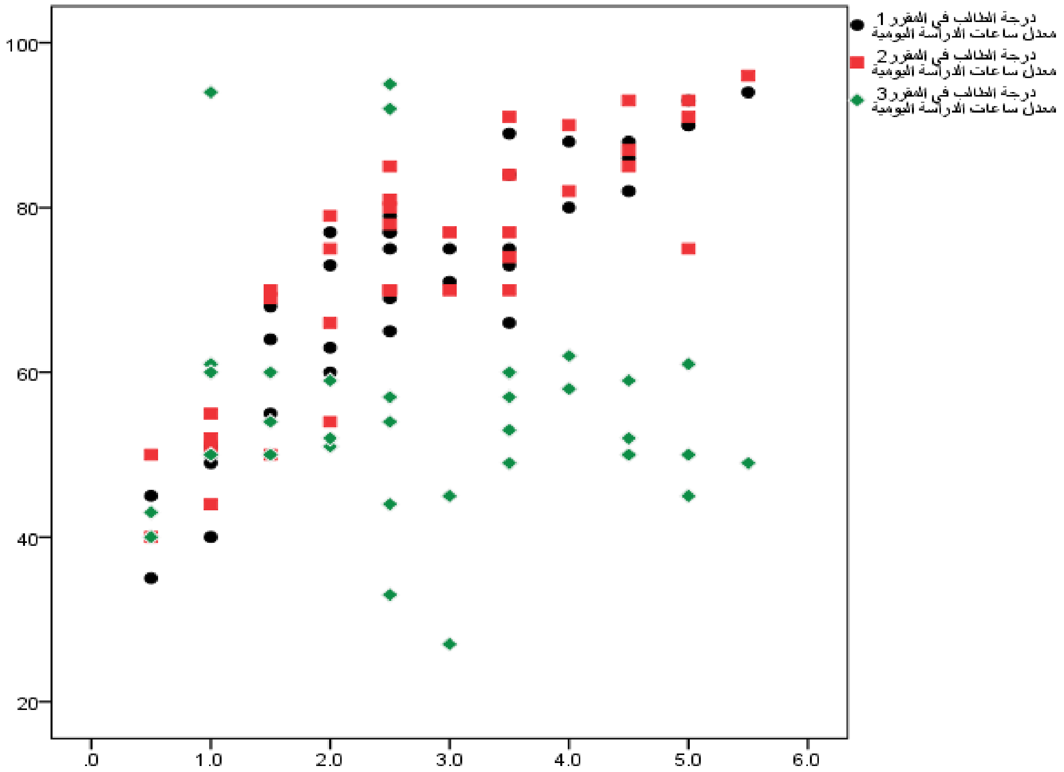


شكل 4.5: نافذة اختيار المتغيرات في شكل الانتشار الفوقي.

في هذه النافذة، سيتم إدخال المتغيرات بصورة أزواج في مربع المحورين السيني والصادي (X-Y Pairs)؛ فيتم إدراج المتغير "مقرر1" أولا ثم المتغير "ساعات_الدراسة"، ثم المتغير "مقرر2" ثم المتغير "ساعات_الدراسة"، ثم المتغير "مقرر3" ثم المتغير "ساعات_الدراسة"، كما هو موضح في الشكل السابق، وبعد الانتهاء من الاختيار نضغط موافق.

سيظهر شكل الانتشار المطلوب في نافذة المخرجات، (مع اختلاف في أشكال وألوان نقاط الرسم بالنسبة للقارئ، حيث أنه تم تغييرها في الشكل للحصول على مزيد من الوضوح في الرؤيا). وننوه هنا إلى أن الشكل (5.5) يمثل

"دمجا" لثلاثة أشكال انتشار هي للمتغيرات "مقرر1" مع "ساعات_الدراسة"، و"مقرر2" مع "ساعات_الدراسة"، و"مقرر3" مع "ساعات_الدراسة" في شكل واحد لتسهيل عملية المقارنة.



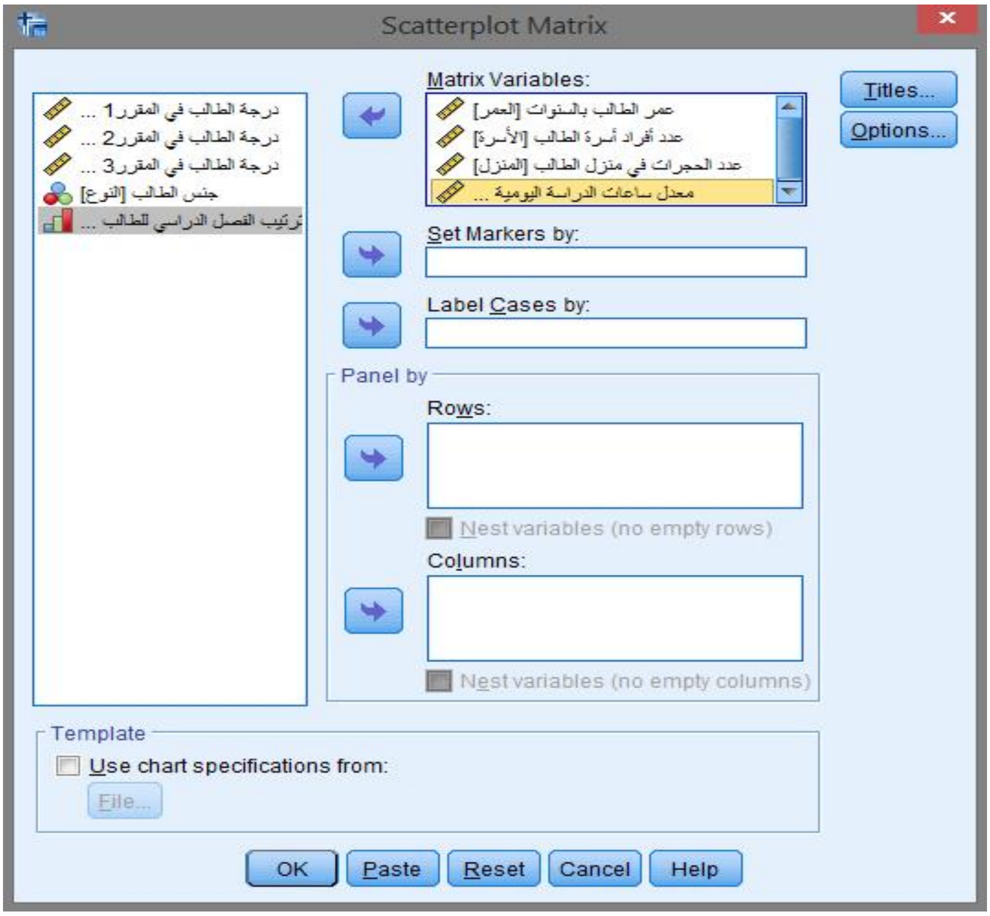
شكل 5.5: شكل الانتشار الفوقي للمتغيرات "مقرر1" و"مقرر2" و"مقرر3" مع "ساعات_الدراسة".

من شكل الانتشار يمكننا ملاحظة وجود علاقة خطية طردية قوية بين معدل ساعات الدراسة ودرجات الطلبة في المقررين 1 و2، أما المقرر 3 فعلاقته ضعيفة جدا مع ساعات الدراسة، أي أن الزيادة في عدد ساعات الدراسة للطلبة يزيد وبشكل كبير من درجات الطلبة في المقررين 1، و2 ولكن لا تزيد درجاتهم في المقرر 3 وهذا قد يدفع الباحث للنظر في طبيعة الظروف المحيطة بهذا المقرر.

نوع آخر من التمثيل البياني لشكل الانتشار هو ما يُعرف **بمصفوفة الانتشار** (Scatterplot Matrix)، والتي تُستخدم لعرض أشكال الانتشار لعدة متغيرات (متغيرين فأكثر) في شكل مصفوفة مربعة كما سنرى؛

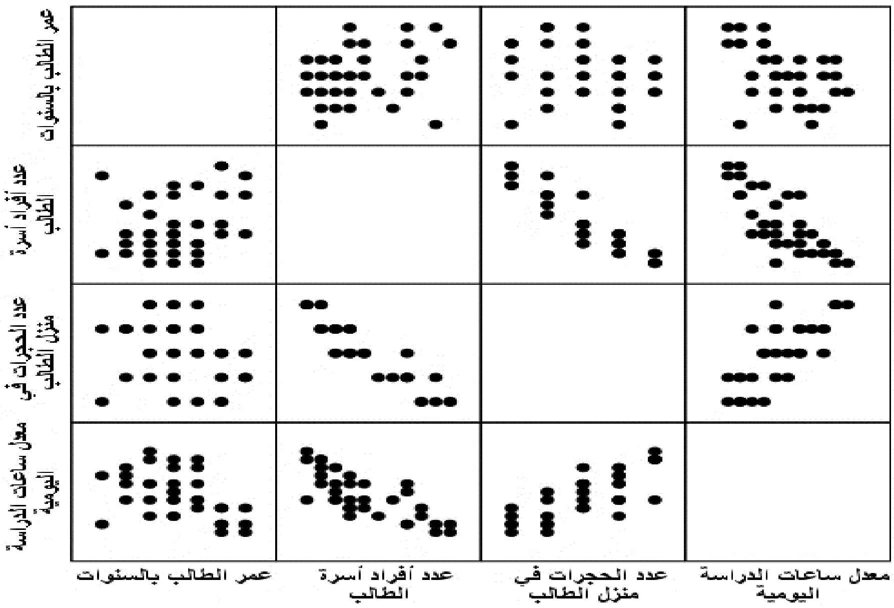
في شريط الأوامر العلوي قم باختيار Graphs>Legacy Dialogs>Scatter/Dot وفي نافذة اختيار شكل الانتشار، (شكل (1.5))، قم باختيار شكل مصفوفة الانتشار (Matrix Scatter) ثم اضغط تعريف فتظهر نافذة جديدة كما في الشكل (6.5).

سنقوم، وعلى سبيل المثال، باختيار المتغيرات "العمر"، "الأسرة"، "المنزل"، و"ساعات_الدراسة" ونقلها إلى مربع متغيرات المصفوفة (Matrix Variables) كما يظهر في الشكل ثم نضغط موافق.



شكل 6.5: نافذة اختيار المتغيرات في مصفوفة الانتشار.

ستظهر مصفوفة شكل الانتشار المطلوبة بعدها في نافذة المخرجات، (شكل (7.5)).



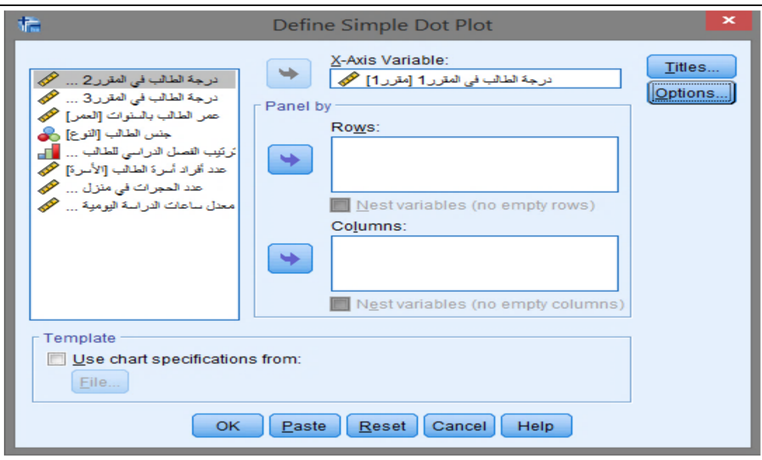
شكل 7.5: مصفوفة الانتشار لمتغيرات "العمر"، "الأسرة"، "المنزل"، و"ساعات_الدراسة".

من الشكل (7.5) يمكننا ملاحظة ما يلي، (حيث أن كل شكل انتشار في المصفوفة هو خاص بالمتغيرين في الصف والعامود للشكل المحدد):

1. وجود علاقة خطية طردية قوية بين معدل ساعات الدراسة اليومية وعدد الحجرات في منزل الطالب، أي أنه بزيادة عدد الحجرات تزيد ساعات الدراسة، وهذا يُعد منطقيا حيث أن المزيد من الحجرات يعني توفر المساحة والمكان الكافي للدراسة.
2. وجود علاقة خطية عكسية قوية بين معدل ساعات الدراسة اليومية وعدد أفراد أسرة الطالب، بمعنى أنه كلما كبر حجم الأسرة كلما أدى ذلك لتناقص في معدل ساعات الدراسة، وربما كان التفسير وراء ذلك هو اكتظاظ منزل الطالب مما يؤثر على الجو العام للمذاكرة.
3. وجود علاقة ضعيفة جدا بين معدل ساعات الدراسة اليومية وعمر الطالب، مما يدل على أن الزيادة أو النقصان في ساعات الدراسة ليس مرتبطا بالعمر.
4. وجود علاقة خطية عكسية قوية بين عدد حجرات المنزل وعدد أفراد أسرة الطالب، وفي ذلك نوع من التناقض المنطقي.
5. وجود علاقة ضعيفة جدا بين عمر الطالب وكلا من عدد حجرات المنزل وعدد أفراد أسرة الطالب، مما يدل على أن الزيادة في عمر الطالب ليست مرتبطة بهاذين المتغيرين.

- **شكل الانتشار لمتغير مفرد (الرسم النقطي) (Dot plot)**

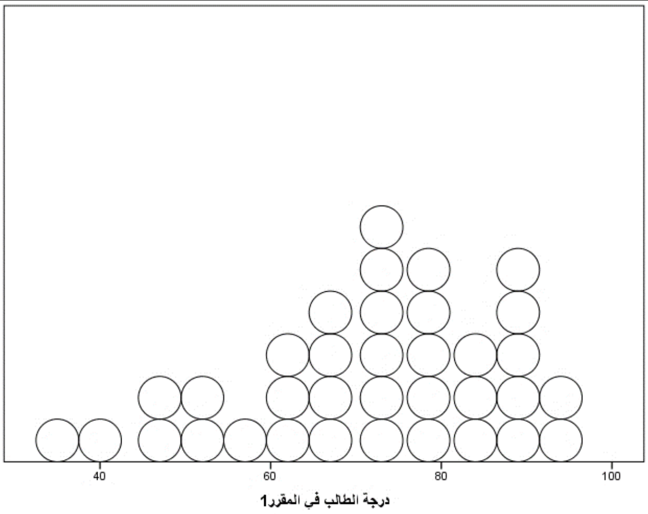
يُعتبر الرسم النقطي من الرسوم البسيطة التي تُستخدم في وصف توزيع متغير واحد على خط الأعداد، ويمكن اعتباره في التفسير مشابها للمدرج التكراري أو شكل الساق والورقة الذي تناولناه سابقا في الفصل الثالث. ولتنفيذ الرسم النقطي، باستخدام ملف البيانات "بيانات الطلبة2"، قم في شريط الأوامر العلوي باختيار Graphs>Legacy Dialogs>Scatter/Dot وفي نافذة اختيار شكل الانتشار، (شكل (1.5))، قم باختيار رسم النقاط البسيط (Simple Dot) ثم اضغط تعريف فتظهر نافذة جديدة كما في الشكل (8.5)؛



شكل 8.5: نافذة تعريف المتغير في الرسم النقطي لملف بيانات "بيانات الطلبة2".

سنختار المتغير "مقرر1" ونقوم بنقله لمربع متغير المحور السيني (X-Axis Variable)، كما هو موضح في الشكل، ويمكن للمستخدم الاختيار من بين ثلاثة طرق عرض للرسم النقطي عند الضغط على زر خيارات (Options)، وهي الرسم النقطي غير المتماثل (Asymmetric) وهو الخيار الافتراضي، الرسم النقطي المتماثل (Symmetric)، والرسم النقطي المسطح (Flat). وفي مثالنا سنبقي على الخيار الافتراضي وهو الرسم النقطي غير المتماثل ونضغط موافق في النافذة الأصلية.

سيظهر الرسم النقطي المطلوب في نافذة المخرجات، (شكل (9.5)). ومن الرسم يمكننا ملاحظة توزيع أو انتشار درجات الطلبة في المقرر 1، والذي يشبه إلى حد كبير المدرج التكراري، إلا أن المدرج التكراري يمثل التكرار المناظر للفترات (في محورين)، أما الرسم النقطي فالنقاط المتراكمة فيه هي تكرار نفس القيمة على محور واحد.



شكل 9.5: الرسم النقطي للمتغير "مقرر1" في ملف بيانات "بيانات الطلبة2".

2.1.5 دراسة الارتباط باستخدام معامل بيرسون

(Analyzing Correlation using Pearson's Coefficient)

إن من أشهر معاملات الارتباط الخطي البسيط التي تستخدم لحساب الارتباط، أي دراسة العلاقة المتبادلة، بين متغيرين كميين هو **معامل بيرسون للارتباط الخطي البسيط** (Pearson Correlation Coefficient)، والذي يعرف للمتغيرين $X = (x_1, x_2, \dots, x_n)$ و $Y = (y_1, y_2, \dots, y_n)$ بالصيغة التالية:

$$r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

وحيث أن $Var(X) = \frac{S_{XX}}{n}$ و $Var(Y) = \frac{S_{YY}}{n}$ و $Cov(X,Y) = \frac{S_{XY}}{n}$، و $-1 \le r_{XY} \le 1$. وهذا يعني أن معامل الارتباط الخطي البسيط ما هو إلا **التغاير** (Covariance) بين المتغيرين X و Y مقسوما على حاصل ضرب كلا من الانحراف المعياري للمتغيرين.

ويتم اختبار معامل الارتباط، (بين المتغيرين X وY)، عن طريق اختبار الفرضية الصفرية؛

$$H_0: \rho_{XY} = 0$$

والتي تعني أن معامل الارتباط للمجتمع ليس له قيمة معنوية، وإحصاءة الاختبار هي الإحصاءة t بـ $(n-2)$ درجات حرية، وتعرف بالصورة:

$$t_c = \frac{r_{XY}}{SE(r_{XY})}$$

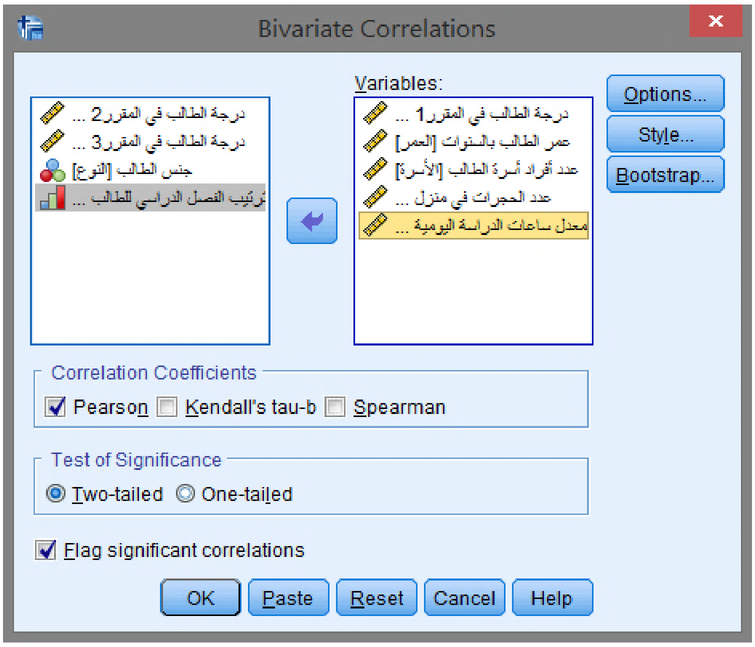
حيث

$$SE(r_{XY}) = \sqrt{\frac{1 - r_{XY}^2}{n-2}}$$

هو الخطأ المعياري لمعامل الارتباط (للعينة). ورفض الفرضية الصفرية عند قيمة احتمالية P-value < 0.05 يدل على أهمية (معنوية) معامل الارتباط أي أهمية العلاقة الخطية بين المتغيرين.

ويمكننا في برنامج SPSS حساب معامل الارتباط البسيط لمتغيرين فقط أو لعدة أزواج من المتغيرات، وهو ما يُعرف بتكوين **مصفوفة الارتباطات البسيطة** (Correlation Matrix)، والتي يُرمز لها في معظم الكتب بالرمز R، ففي ملف البيانات "بيانات الطلبة2" يمكننا حساب معاملات الارتباط الخطي البسيط بالطريقة التالية:

في شريط الأوامر العلوي قم باختيار Analyze>Correlate>Bivariate فتظهر نافذة الارتباط الثنائي أو البسيط (Bivariate Correlations) كما هو في الشكل (10.5)؛



شكل 10.5: نافذة حساب الارتباطات البسيطة، (ملف بيانات "بيانات الطلبة2").

سنقوم على سبيل المثال بحساب معاملات الارتباط الخطي البسيط بين المتغيرات "مقرر1"، "العمر"، "الأسرة"، "المنزل"، و"ساعات_الدراسة". وهكذا نقوم بنقل هذه المتغيرات إلى مربع المتغيرات (Variables) على اليمين. ونلاحظ في منتصف النافذة وجود ثلاثة خيارات أو ثلاثة أنواع من معاملات الارتباط البسيط وهي معامل بيرسون، معامل كندل تاو–B، ومعامل ارتباط الرتب لسبيرمان (Spearman)، والمعاملين الأخيرين يُستخدمان عادة لحساب الارتباط البسيط بين المتغيرات الوصفية، (كما رأينا سابقا مع معمل كندل في جداول الاقتران في الفصل السابق).

وحيث أن كل المتغيرات المختارة هنا هي متغيرات كمية فسنختار حساب معامل بيرسون فقط، وهو الخيار الافتراضي للبرنامج. وسنبقي أيضا على خيار الاختبار من طرفين (Two-tailed) الافتراضي، ونبقي أيضا على خيار تمييز الارتباطات المعنوية (Flag significant correlations)، ثم نضغط موافق فنحصل على مصفوفة الارتباطات البسيطة بين المتغيرات المُختارة كما يظهر في جدول (2.5).

جدول 2.5: مصفوفة الارتباطات بين المتغيرات "مقرر1"، "العمر"، "الأسرة"، "المنزل"، و"ساعات_الدراسة"، ("بيانات الطلبة2").

Correlations

		درجة الطالب في المقرر 1	عمر الطالب بالسنوات	عدد أفراد أسرة الطالب	عدد الحجرات في منزل الطالب	معدل ساعات الدراسة اليومية
درجة الطالب في المقرر 1	Pearson Correlation	1	-.416-*	-.904-**	.865**	.897**
	Sig. (2-tailed)		**.013**	**.000**	**.000**	**.000**
	N	35	35	35	35	35
عمر الطالب بالسنوات	Pearson Correlation	-.416-*	1	.292	-.273	-.456-**
	Sig. (2-tailed)	.013		**.089**	**.112**	**.006**
	N	35	35	35	35	35
عدد أفراد أسرة الطالب	Pearson Correlation	-.904-**	.292	1	-.926-**	-.764-**
	Sig. (2-tailed)	.000	.089		**.000**	**.000**
	N	35	35	35	35	35
عدد الحجرات في منزل الطالب	Pearson Correlation	.865**	-.273	-.926-**	1	.756**
	Sig. (2-tailed)	.000	.112	.000		**.000**
	N	35	35	35	35	35
معدل ساعات الدراسة اليومية	Pearson Correlation	.897**	-.456-**	-.764-**	.756**	1
	Sig. (2-tailed)	.000	.006	.000	.000	
	N	35	35	35	35	35

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

وفي هذه المصفوفة المربعة المتماثلة، يكون لدينا ثلاثة قيم في كل "خلية" في الجدول وهي قيمة معامل بيرسون (Pearson Correlation) والقيمة الاحتمالية (.Sig) وحجم العينة (N). ونستطيع ملاحظة العلاقات التالية:

1. وجود علاقة خطية عكسية قوية بين درجة الطالب في المقرر 1 وكل من عمر الطالب (P-value = 0.013) وعدد أفراد الأسرة (P-value = 0)، أي أن درجة الطالب في المقرر 1 تنخفض بتقدم عمر الطالب وزيادة عدد أفراد أسرته، والعلاقة الأولى قد تكون متناقضة مع الواقع العملي، إذ أن زيادة عمر الطالب من المفترض أن تصاحبها زيادة في المستوى الدراسي بسبب زيادة الخبرة الأكاديمية عنده واقترابه من مرحلة التخرج.
2. وجود علاقة خطية طردية قوية بين درجة الطالب في المقرر 1 وكل من عدد حجرات منزل الطالب ومعدل ساعات الدراسة اليومية، بمعنى أن درجة الطالب في المقرر 1 تزيد بزيادة كل من عدد الحجرات في المنزل ومعدل ساعات الدراسة اليومية، وهذه العلاقات منطقية جدا.
3. توجد علاقة طردية ضعيفة (غير معنوية) بين عمر الطالب وعدد أفراد الأسرة، وعلاقة عكسية ضعيفة بين العمر وعدد حجرات المنزل، وهذا يُعد منطقيا بالنسبة لمتغير العمر. إلا أن علاقة العمر بمعدل ساعات الدراسة اليومية يُعد معنويا لوجود علاقة عكسية قوية تعكس انخفاض ساعات الدراسة للطالب بزيادة عمره.
4. يرتبط زيادة عدد أفراد أسرة الطالب بانخفاض عدد حجرات المنزل، (علاقة عكسية قوية)، وهذه العلاقة ربما تكون نتيجة عدم تمكن الكثير من الأسر من الانتقال لمنزل أكبر عند زيادة عدد أفرادها. ويرتبط عدد أفراد أسرة الطالب أيضا بانخفاض معدل ساعات الدراسة للطالب، وهذا يعتبر منطقيا لأن العدد الكبير لأفراد الأسرة عادة ما يسبب اكتظاظ المنزل وعدم حصول الطالب على المساحة الكافية للمذاكرة.
5. وجود علاقة طردية قوية بين عدد حجرات منزل الطالب ومعدل ساعات الدراسة اليومية، وهذا طبيعي حيث أن توفر المساحة المناسبة للمذاكرة يشجع الطالب عادة على زيادة ساعات الدراسة اليومية.

3.1.5 **معامل الارتباط الجزئي** (Partial Correlation Coefficient)

في بعض الدراسات التي تتطلب فهم العلاقة المتبادلة ضمن عدة متغيرات، قد نكون مهتمين بتحليل العلاقة بين ثلاثة أو أربعة متغيرات بحيث يتم حساب الارتباط بين متغيرين منهم مع "تثبيت" أو "وجود" أثر متغير ثالث أو متغيرين، وهو ما يعرف بحساب معامل الارتباط الجزئي.

لنفرض أنه تم حساب معاملات الارتباط الخطي البسيط لأربعة متغيرات هي X_1، X_2، X_3، وX_4، وتم عرضها في مصفوفة الارتباط المتماثلة التالية:

$$R = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} \\ & 1 & r_{23} & r_{24} \\ & & 1 & r_{34} \\ & & & 1 \end{bmatrix} \end{array}$$

عندئذ يمكننا تعريف بعض معاملات الارتباط الجزئية مثل معامل الارتباط الجزئي للمتغيرين X_1 و X_2 مع وجود المتغير X_3 بالصيغة التالية؛

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

حيث $-1\le r_{12.3}\le 1$. وبالمثل، يمكن حساب $r_{13.2}$ و $r_{23.1}$ بالصورة $r_{13.2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}}$ و $r_{23.1}=\frac{r_{23}-r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}}$.

أما صيغة معامل الارتباط الجزئي للمتغيرين X_1 و X_2 بتثبيت المتغيرين X_3 و X_4 فهي تعطى بالصيغة:

$$r_{12.34}=\frac{r_{12.4}-r_{13.4}r_{23.4}}{\sqrt{(1-r_{13.4}^2)(1-r_{23.4}^2)}}=\frac{r_{12.3}-r_{14.3}r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}}$$

وبالمثل يمكن حساب باقي معاملات الارتباط الجزئي الأخرى.

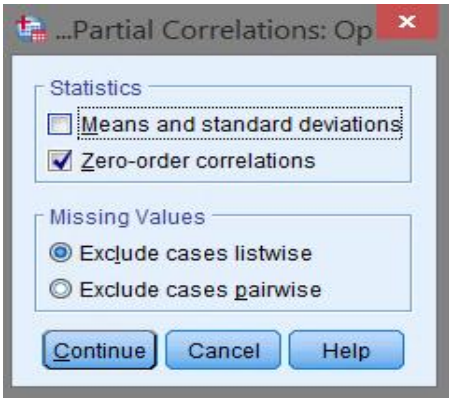
وسنقوم الآن بحساب معامل الارتباط الجزئي لبعض المتغيرات في ملف البيانات "بيانات الطلبة2" في SPSS. سنقوم في شريط الأدوات العلوي باختيار Analyze>Correlate>Partial فتظهر النافذة التالية، (شكل (11.5))؛



شكل 11.5: نافذة حساب الارتباط الجزئي لملف بيانات "بيانات الطلبة2".

سنقوم على سبيل المثال بحساب معامل الارتباط الجزئي بين المتغيرين "المقرر1" و"ساعات_الدراسة" بوجود المتغير "العمر". فنقوم في تلك النافذة بنقل المتغيرين الأولين إلى مربع المتغيرات (Variables) ونقل المتغير الثالث إلى مربع متغير التحكم (Controlling for) كما يوضح الشكل (11.5).

ولمقارنة معامل الارتباط الجزئي بمعامل الارتباط البسيط لنفس المتغيرين، (حيث أن ذلك سيعطي معلومات حول مدى تأثير المتغير الثالث على العلاقة بين المتغيرين الأساسيين)، نضغط على زر الخيارات (Options) في نافذة الارتباط الجزئي فتظهر نافذة فرعية كما هو في الشكل (12.5).



شكل 12.5: نافذة الخيارات للارتباط الجزئي.

في هذه النافذة الفرعية، قم باختيار عرض الارتباطات البسيطة (Zero-order correlations) كما هو موضح ثم اضغط استمرار للعودة لنافذة الارتباط الجزئي. وفي تلك النافذة اضغط موافق للحصول على معامل الارتباط الجزئي المطلوب مصحوبا بمعاملات الارتباط البسيط بين كل متغيرين.

في نافذة المخرجات، سنحصل على معامل الارتباط الجزئي بين المتغيرين "المقرر1" و"ساعات_الدراسة" بوجود المتغير "العمر"، (جدول (3.5))، والذي يُظهر وجود علاقة طردية قوية جدا $\left(r_{(\text{المقرر1})(\text{ساعات_الدراسة}).(\text{العمر})} = 0.874\right)$ بين درجات الطلبة في مقرر 1 ومعدل ساعات دراستهم اليومية بوجود تأثير متغير العمر.

جدول 3.5: معامل الارتباط الجزئي بين المتغيرين "المقرر1" و"ساعات_الدراسة" بوجود متغير "العمر" مع معاملات الارتباط البسيط للمتغيرات الثلاثة، (ملف بيانات "بيانات الطلبة2").

Correlations

Control Variables			درجة الطالب في المقرر1	معدل ساعات الدراسة اليومية	عمر الطالب بالسنوات
-none-[a]	درجة الطالب في المقرر1	Correlation	1.000	**.897**	-.416
		Significance (2-tailed)		.000	.013
		df	0	33	33
	معدل ساعات الدراسة اليومية	Correlation	.897	1.000	-.456
		Significance (2-tailed)	.000		.006
		df	33	0	33
	عمر الطالب بالسنوات	Correlation	-.416	-.456	1.000
		Significance (2-tailed)	.013	.006	
		df	33	33	0
عمر الطالب بالسنوات	درجة الطالب في المقرر1	Correlation	1.000	**.874**	
		Significance (2-tailed)		**.000**	
		df	0	32	
	معدل ساعات الدراسة اليومية	Correlation	.874	1.000	
		Significance (2-tailed)	.000		
		df	32	0	

a. Cells contain zero-order (Pearson) correlations.

ولكن إذا ما تم مقارنة قيمة معامل الارتباط الجزئي بقيمة معامل الارتباط البسيط بين درجات الطلبة في المقرر1 ومعدل ساعات دراستهم اليومية، $\left(r_{(\text{ساعات_الدراسة})(\text{المقرر1})} = 0.897 \right)$، حيث أن قيم المعاملين متقاربة جدا، فإننا

نستنتج أن المتغير "العمر" ليس له أهمية معنوية في علاقة المتغيرين "المقرر1" و"ساعات_الدراسة" ببعضهما البعض.

من جديد لنقم بحساب معامل الارتباط الجزئي بين المتغيرين "المقرر2" و"ساعات_الدراسة" بوجود المتغير "الأسرة"، وبتطبيق نفس الخطوات السابقة، نحصل على النتيجة في نافذة المخرجات، (جدول (4.5)).

جدول 4.5: معامل الارتباط الجزئي بين المتغيرين "المقرر2" و"ساعات_الدراسة" بوجود متغير "الأسرة" مع معاملات الارتباط البسيط للمتغيرات الثلاثة، (ملف بيانات "بيانات الطلبة2").

Correlations

Control Variables			معدل ساعات الدراسة اليومية	درجة الطالب في المقرر2	عدد أفراد أسرة الطالب
-none-[a]	معدل ساعات الدراسة اليومية	Correlation	1.000	**.847**	-.764
		Significance (2-tailed)		.000	.000
		df	0	33	33
	درجة الطالب في المقرر2	Correlation	.847	1.000	-.907
		Significance (2-tailed)	.000		.000
		df	33	0	33
	عدد أفراد أسرة الطالب	Correlation	-.764	-.907	1.000
		Significance (2-tailed)	.000	.000	
		df	33	33	0
عدد أفراد أسرة الطالب	معدل ساعات الدراسة اليومية	Correlation	1.000	**.566**	
		Significance (2-tailed)		**.000**	
		df	0	32	
	درجة الطالب في المقرر2	Correlation	.566	1.000	
		Significance (2-tailed)	.000		
		df	32	0	

a. Cells contain zero-order (Pearson) correlations.

وبمقارنة قيمة معامل الارتباط الجزئي بقيمة معامل الارتباط البسيط بين درجات الطلبة في المقرر2 ومعدل ساعات دراستهم اليومية نرى بأن $r_{(\text{المقرر 2})(\text{ساعات_الدراسة}).(\text{الأسرة})} < r_{(\text{المقرر 2})(\text{ساعات_الدراسة})}$ ، وهذا يدل على أن وجود متغير عدد أفراد الأسرة قد أدى إلى انخفاض حاد في العلاقة الطردية بين المتغيرين الأساسيين ("المقرر2" و"ساعات_الدراسة")، أي أن زيادة عدد أفراد يؤثر سلبا في العلاقة بين درجات الطلبة في المقرر2 ومعدل ساعات الدراسة. وبنفس الطريقة يمكن دراسة العلاقة بين أي متغيرين بوجود متغير أو أكثر[1].

2.5 الارتباط بين المتغيرات الوصفية والكمية

(Correlation between Categorical and Quantitative Variables)

في الفصل السابق، تطرقنا لدراسة العلاقة بين متغيرين وصفيين يمثلان مستويات محددة في جدول الاقتران واختبار هذه العلاقة، وفي هذا البند سنتناول كيفية حساب معامل الارتباط بين متغيران وصفيان بصورة عامة، وكيفية تحليل العلاقات بين المتغيرات الكمية والوصفية.

[1] يمكن إضافة أكثر من متغير ثانوي في مربع متغير التحكم (Controlling for) كما هو موضح في الشكل (11.5).

باستخدام ملف البيانات "فيتامين دال"، سنقوم بحساب معاملات الارتباط بين متغيرات هذه البيانات بالصورة التالية؛ في شريط الأوامر العلوي قم باختيار Analyze>Correlate>Bivariate فتظهر نافذة الارتباط الثنائي (Bivariate Correlations) كما رأينا في الشكل (10.5). قم باختيار المتغيرات "النوع"، "الفئة_العمرية"، و"مستويات_فيتامين_دال" ونقلها إلى مربع المتغيرات (Variables).

وفي منتصف النافذة، قم باختيار الأساليب الثلاثة لمعاملات الارتباط وهي معامل بيرسون، معامل كندل تاو−B، ومعامل ارتباط الرتب لسبيرمان، وقد قمنا هنا باختيار معامل بيرسون[1]، رغم أنه مخصص لحساب الارتباط بين المتغيرات الكمية، بهدف مقارنة نتائجه بنتائج معاملات الارتباط للمتغيرات الوصفية؛ معامل كندل تاو−B، ومعامل سبيرمان. بعد ذلك اضغط موافق فتظهر النتائج في نافذة المخرجات. النتيجة الأولى، الموضحة في جدول (5.5)، هي مصفوفة الارتباطات للمتغيرات الثلاثة التي تم تحديدها، وهي تُظهر علاقات ارتباط عكسية قوية ذات معنوية بين مستويات فيتامين دال وكل من النوع وفئات العمر، وأيضا علاقة عكسية ضعيفة بين فئات العمر والنوع.

جدول 5.5: مصفوفة معاملات الارتباط لبيرسون بين المتغيرات "النوع"، "الفئة_العمرية"، و"مستويات_فيتامين_دال".

Correlations

		النوع	فئات العمر	مستويات فيتامين دال
النوع	Pearson Correlation	1	**-.201**	**-.499-******
	Sig. (2-tailed)		**.123**	**.000**
	N	60	60	60
فئات العمر	Pearson Correlation	-.201	1	**-.367-******
	Sig. (2-tailed)	.123		**.004**
	N	60	60	60
مستويات فيتامين دال	Pearson Correlation	-.499-**	-.367-**	1
	Sig. (2-tailed)	.000	.004	
	N	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

وأما النتيجة الثانية، (جدول (6.5))، فهي مصفوفة الارتباطات للمتغيرات الثلاثة، باستخدام معامل كندل أولا ومعامل سبيرمان ثانيا، وكلا المعاملين يُظهر علاقات ارتباط قريبة جدا من تلك المحسوبة بمعامل بيرسون في الجدول (5.5). والاستنتاج سيكون كالتالي:

1. العلاقة العكسية القوية بين مستويات فيتامين دال وفئات العمر، تعني أنه كلما زادت الفئة العمرية للشخص كلما نقصت لديه معدلات فيتامين دال، وهذه النتيجة مطابقة لما سبق.
2. أما العلاقة العكسية القوية بين مستويات فيتامين دال والنوع فلا يتم التعليق عليها كالنتيجة السابقة حيث أن النوع هو متغير وصفي اسمي وليس رتبي، ونلفت انتباه القارئ هنا إلى أنه لن يكون هنالك معنى أن نقول "كلما نقص النوع كلما زادت معدلات فيتامين دال". ويمكن في هذه الحالة دراسة الارتباط بين متغير "النوع" ومتغير كمي آخر باستخدام شكل الانتشار كما سنوضح تاليا، وكذلك الأمر بالنسبة للعلاقة بين فئات العمر والنوع.

[1] لابد من التأكيد هنا على أن معامل بيرسون يستخدم، من الناحية النظرية، لحساب الارتباط بين المتغيرات الكمية التي تتبع التوزيع الطبيعي.

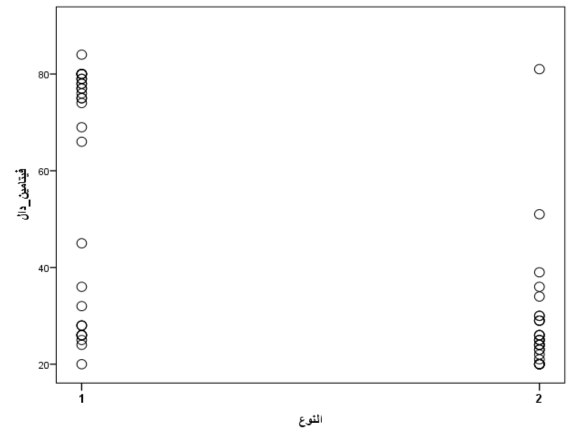
جدول 6.5: مصفوفة معاملات الارتباط لكندل وسبيرمان بين متغيرات "النوع"، "الفئة_العمرية"، و"مستويات_فيتامين_دال".

Correlations

			النوع	فئات العمر	مستويات فيتامين دال
Kendall's tau_b	النوع	Correlation Coefficient	1.000	-.190	-.466-**
		Sig. (2-tailed)		.122	.000
		N	60	60	60
	فئات العمر	Correlation Coefficient	-.190	1.000	-.347-**
		Sig. (2-tailed)	.122		.003
		N	60	60	60
	مستويات فيتامين دال	Correlation Coefficient	-.466-**	-.347-**	1.000
		Sig. (2-tailed)	.000	.003	
		N	60	60	60
Spearman's rho	النوع	Correlation Coefficient	1.000	-.201	-.490-**
		Sig. (2-tailed)		.123	.000
		N	60	60	60
	فئات العمر	Correlation Coefficient	-.201	1.000	-.372-**
		Sig. (2-tailed)	.123		.003
		N	60	60	60
	مستويات فيتامين دال	Correlation Coefficient	-.490-**	-.372-**	1.000
		Sig. (2-tailed)	.000	.003	
		N	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

لتوضيح الملاحظة السابقة لنقم بالخطوات التالية؛ في شريط الأوامر العلوي اختر Graphs>Legacy Dialogs>Scatter/Dot ثم اختر الانتشار البسيط (Simple Scatter) من نافذة الرسم، (شكل (1.5))، ثم اضغط تعريف. في نافذة تعريف المتغيرات، (شكل (2.5))، اختر المتغير الكمي "فيتامين_دال" وقم بنقله لمربع المحور الصادي (Y-Axis) واختر متغير "النوع" وانقله لمربع المحور السيني (X-Axis) ثم اضغط موافق.



شكل 13.5: شكل الانتشار للمتغيرين "فيتامين_دال" و"النوع".

في نافذة المخرجات سيظهر شكل الانتشار التالي، (شكل[1] (13.5))، ومنه يمكننا ملاحظة أن معدلات فيتامين دال عند النساء (حيث أن الرقم 2 في المحور الأفقي يشير للإناث) هي في العموم أقل من الرجال (رقم 1 في المحور الأفقي) في مختلف الأعمار.

[1] ننوه هنا إلى أننا قمنا ببعض التعديلات على الرسم فيما يخص القيم في المحور الأفقي وألوان النقاط في الرسم.

3.5 تحليل الانحدار الخطي البسيط (Simple Linear Regression Analysis)

في البندين السابقين، تناولنا مفهوم الارتباط بين المتغيرات الكمية/الكمية، والوصفية/الوصفية، والكمية/الوصفية، ووضحنا كيفية حساب هذه المعاملات جبريا من خلال قوانين الارتباط، وبصريا من خلال انتشار نقاط البيانات، وناقشنا كيفية استنتاج مدى قوة ونوع العلاقة بين المتغيرات.

إلا أن دراسة الارتباط بمفرده قد لا يكون كافيا في كثير من الحالات لأنه لا يتعرض **للعلاقة السببية** (Causal Relationship) بين المتغيرات؛ ويقصد بالعلاقة السببية هنا كيفية وقوة تأثير متغير ما (أو أكثر) على متغير آخر (أو أكثر)، فمثلا؛ يمكن استخدام معامل الارتباط لدراسة قوة ونوع العلاقة بين سعر سلعة ما (X) وعدد الوحدات المباعة منها (Y)، والتي نتوقع منطقيا أن تكون علاقة عكسية قوية، إلا أننا لا نستطيع معرفة مدى تأثير سعر السلعة على عدد الوحدات المباعة باستخدام معامل الارتباط فقط، لذلك نحن بحاجة لمقياس أو أسلوب إضافي لدراسة ذلك التأثير، وهنا يأتي دور تحليل الانحدار للقيام بهذه المهمة، وبالتالي يمكننا القول أنه لدراسة العلاقات بين المتغيرات بصورة أكثر تعمقا لابد لنا من استخدام كلا من الأسلوبين؛ تحليل الانحدار والارتباط.

إضافة إلى ذلك، يتم استخدام أسلوب تحليل الانحدار لغرض **التنبؤ** (Prediction)، والذي يعد من أهم أهداف الأسلوب، فمثلا؛ إذا وجد أن هنالك تأثير معنوي لسعر السلعة على عدد الوحدات المباعة، بمعنى أن عدد الوحدات المباعة يعتمد على سعر السلعة، فيمكن عندها التنبؤ بعدد الوحدات المباعة (كتقدير مستقبلي أو كدراسة جدوى) عند ارتفاع أو انخفاض سعر السلعة، أي عند أي قيمة افتراضية للسعر.

وعمليا يمكن في الدراسات الاستكشافية استخدام تحليل الارتباط أولا لتبيان قوة العلاقة بين المتغيرات، فإن كان هنالك علاقة معنوية بينها فيمكن عندئذ المضي قدما باستخدام تحليل الانحدار لدراسة العلاقة السببية. وكما هو الحال مع تحليل الارتباط، فإن أسلوب تحليل الانحدار يتعامل مع المتغيرات التي توجد بينها علاقات خطية وغير خطية، إلا أننا سنركز على الانحدار الخطي فقط كما ذكرنا سابقا.

1.3.5 تعريف نموذج الانحدار الخطي البسيط (Definition of Simple Linear Regression Model)

إذا كان X وY متغيران كميان فإن نموذج الانحدار الخطي لهما يعرف بالصورة:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , i = 1, 2, \dots, n$$

حيث:

Y : يعرف **بالمتغير التابع** أو **متغير الاستجابة** (Dependent or Response Variable)، X : يعرف **بالمتغير التوضيحي** أو **المستقل** أو **متغير التنبؤ** (Explanatory or Independent Variable or Predictor)،

β_0 و β_1 تعرف بمعالم النموذج، (ويسمى β_1 بمعامل الانحدار)، ε يعرف بحد الخطأ العشوائي للنموذج، و n هو عدد المشاهدات أو حجم عينة الدراسة.

وحد الخطأ العشوائي تم إضافته لنموذج الانحدار لعدة أسباب أهمها أن أي عملية تقدير إحصائي لابد لها من هامش خطأ للتقدير لأن القيمة المقدرة للمعلمة، (بصورة عامة)، لن تساوي القيمة الحقيقة للمعلمة كما تنص قواعد التقدير. إلا أننا نطمح دائما لتقديرات تكون قريبة أكثر ما يمكن للقيمة الفعلية.

ويتم تقدير معالم النموذج بالصيغ:

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{Cov(X,Y)}{Var(X)} \text{ ، } \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$

حيث $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ و $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. ومن الصيغة السابقة، يمكننا ملاحظة أن العلاقة بين معامل الارتباط الخطي ومعامل الانحدار الخطي تكون بالصورة؛

$$\hat{\beta}_1 = \frac{\sqrt{S_{YY}}}{\sqrt{S_{XX}}} r_{XY} = \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}} r_{XY}$$

2.3.5 فرضيات طريقة المربعات الصغرى الاعتيادية (Assumptions of OLS Method)

إن استخدام طريقة المربعات الصغرى الاعتيادية ((Ordinary Least Squares, (OLS)) في التقدير يكون من الناحية النظرية مقيدا بتحقق مجموعة من الشروط أو الفرضيات، هذه الفرضيات تسمى أيضا فرضيات نموذج الانحدار الخطي. ومن الناحية العملية، فإنه عندما يقوم الباحث بتقدير معالم نموذج الانحدار فإنه لا يقوم عادة بالتحقق من صحة هذه الفرضيات للنموذج الموفق، وهذا يُعد "نقصا" في نتائج تحليل الانحدار لأن ثبات أو تحقق هذه الفرضيات يدل على ثبات أو جودة النتائج التي سيتم الوصول إليها عن طريق النموذج المقدر.

وفيما يلي سنقوم بتلخيص أهم هذه الفرضيات، مع التذكير بأن هذه الفرضيات تسري على كل من النماذج الخطية البسيطة، (التي تضم متغير توضيحي واحد)، والمتعددة، (التي تضم أكثر من متغير توضيحي)؛

لتوفيق نموذج الانحدار الخطي: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , i = 1, 2, \dots, n$ باستخدام طريقة المربعات الصغرى الاعتيادية لابد من توفر الفرضيات التالية:

1. **خطية النموذج**، والتي تعني أن المتغير التابع هو تركيبة خطية من المتغير التوضيحي (أو المتغيرات التوضيحية) ومعالم النموذج.
2. المتغير التوضيحي (أو المتغيرات التوضيحية) تأخذ **قيما غير عشوائية** (Non-Random)، بمعنى أنها قيما لا تتبع توزيعا احتماليا.

3. **المتغيرات التوضيحية** (في نماذج الانحدار المتعدد) لابد أن تكون جميعها **مستقلة خطيا** عن بعضها البعض، ويُقصد بذلك ألا يوجد بين أي متغيرين علاقة خطية (أي أن يكون معامل الارتباط الخطي بينهما مساو للصفر).

.4

أ- أن تكون **قيم حد الخطأ العشوائي مستقلة** عن بعضها البعض، (بمعنى أن الارتباط أو التغاير بينها مساو للصفر؛ $Cov(\varepsilon_i, \varepsilon_j) = 0,\ \forall i \neq j = 1, 2, \dots, n$).

ب- **حد الخطأ يتبع توزيعا طبيعيا** بمتوسط يساوي 0 وتباين يساوي σ_ε^2، أي أن $E(\varepsilon_i) = 0$، و $Var(\varepsilon_i) = \sigma_\varepsilon^2$، لكل $i = 1, 2, \dots, n$.

ج- **تجانس تباين قيم حد الخطأ**، بمعنى أن يكون للأخطاء نفس قيمة التباين مهما تغيرت قيم المتغير التابع، ($Var(\varepsilon_i) = \sigma_\varepsilon^2, \forall i = 1, 2, \dots, n$).

ويمكن تلخيص الفرضيات (أ، ب، و ج) بالصورة؛

$$\varepsilon_i \sim NID(0, \sigma_\varepsilon^2), \forall i = 1, 2, \dots, n$$

3.3.5 التمثيل البياني لنموذج الانحدار الخطي البسيط

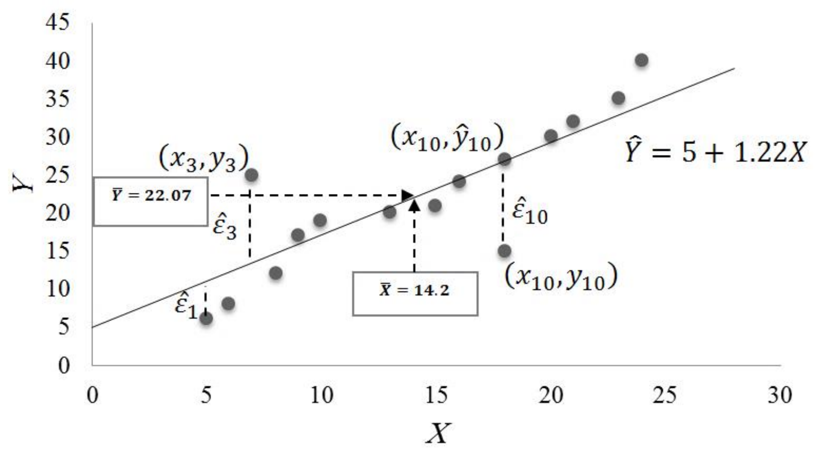
(Graphical Representation for the Linear Regression Model)

بعد تقدير معالم نموذج الانحدار الخطي β_0 و β_1 ، وهو ما يعرف **بتوفيق النموذج** (Model Fitting)، يمكن تمثيل هذا النموذج أو رسم خط الانحدار بيانيا مباشرة على شكل الانتشار للمتغيرين وذلك من خلال تحديد قيم الجزء المقطوع من المحور الصادي $\hat{\beta}_0$ و الميل $\hat{\beta}_1$ على الرسم أو من خلال التعويض عن أي قيمتين للمتغير X في نموذج الانحدار وتمرير خط مستقيم عبرهما.

والشكل التالي (شكل (14.5))، يوضح شكل الانتشار لبيانات افتراضية وخط الانحدار الذي يمر عبرها وأيضا يُظهر تقديرات قيم حد الخطأ ε التي تسمى **بالبواقي** $\hat{\varepsilon}$ (Residuals)، وتمثل الفرق بين قيم المتغير التابع الفعلية وقيمه المقدرة، أي أنها تحسب بالصيغة؛

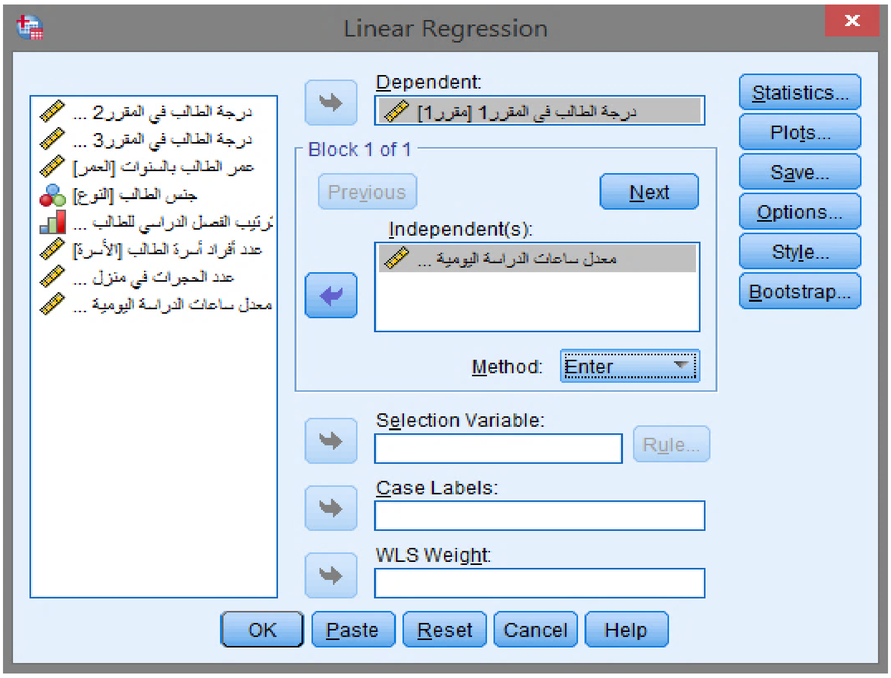
$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad , i = 1, 2, \dots, n$$

مما يعني أنه كلما قلت قيمة البواقي كلما دل ذلك على "جودة" توفيق النموذج وملائمته لبيانات العينة، ومن خواص هذه البواقي أن بعض قيمها ستكون سالبة والأخرى موجبة، وأن مجموع تلك القيم لابد أن يساوي الصفر، أي أن $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$. ولاحظ أيضا من الشكل أن قيم متوسطات المتغيرين X وY لابد أن تشكل إحدى نقاط خط الانحدار دائما، والتي كلما اقتربت من منتصف كتلة البيانات كلما دل ذلك على قلة القيم المتطرفة.



شكل 14.5: شكل الانتشار وتوفيق خط الانحدار لبيانات افتراضية.

ولنقم الآن بتوفيق نموذج انحدار خطي بسيط باستخدام بيانات المثال "بيانات الطلبة2". سنقوم في شريط الأوامر العلوي باختيار Analyze>Regression>Linear فتظهر نافذة تحليل الانحدار الخطي كما يظهر في الشكل (15.5). وعند اختيار المتغير التابع والمتغير أو المتغيرات التوضيحية، يجب أن يراعي الباحث العلاقة المنطقية السببية بين المتغيرات، أي أنه يجب أن يختار المتغير التابع الذي "يعتقد" أنه يعتمد أو يتأثر بتغير المتغير التوضيحي.

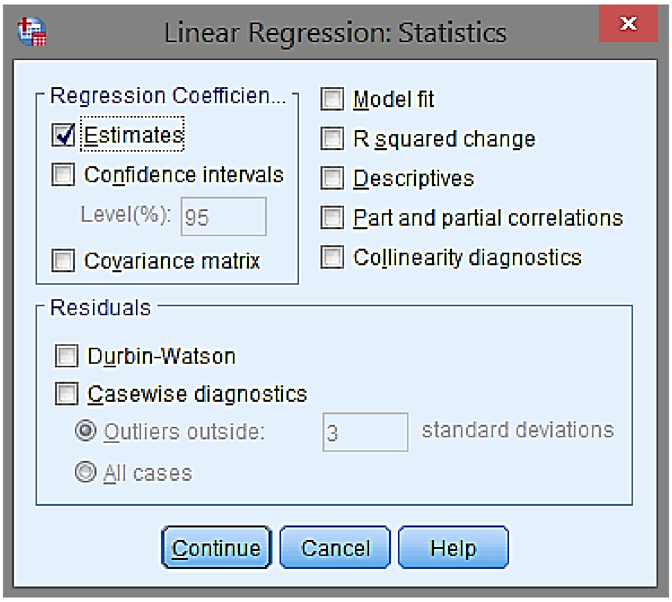


شكل 15.5: نافذة تحليل الانحدار الخطي لملف البيانات "بيانات الطلبة2".

فعلى سبيل المثال، واعتمادا على منطق السببية، يمكننا اختيار درجات الطلبة في المقرر 1 كمتغير استجابة يتأثر بتغير معدل ساعات الدراسة اليومية للطلبة، لذلك سنقوم باختيار المتغير "المقرر 1" ونقله إلى مربع المتغير التابع

(Dependent)، واختيار المتغير "ساعات_الدراسة" ونقله لمربع المتغيرات التوضيحية أو المستقلة (Independent)، كما يوضح الشكل (15.5).

في الخطوة التالية، سنقوم بتوفيق نموذج انحدار المتغير التابع "المقرر 1" على المتغير التوضيحي "ساعات_الدراسة" مستخدمين الحد الأدنى للنتائج، بمعنى أننا سنقوم باختيار عرض أقل عدد من النتائج بهدف التدرج مع القارئ في شرح كيفية تفسير النتائج خطوة بخطوة.



شكل 16.5: نافذة خيارات عرض الإحصاءات في تحليل الانحدار.

في نافذة الانحدار الخطي، (الشكل (15.5))، اضغط على خيار الإحصاءات (Statistics) فتظهر النافذة الفرعية التالية، (شكل (16.5))، في تلك النافذة قم باختيار عرض التقديرات (Estimates) فقط ثم اضغط استمرار للعودة للنافذة الأصلية. اضغط بعد ذلك موافق فتظهر النتائج في نافذة المخرجات.

الجدول الأول في النتائج يوضح الطريقة المتبعة في إدخال المتغيرات في النموذج، حيث أنه توجد عدة طرق "لبناء" نماذج الانحدار سنتطرق لها لاحقا في هذا الفصل، أما الجدول التالي، (جدول (7.5))، فيتم فيه عرض نتائج توفيق نموذج الانحدار الخطي المطلوب، ومن هذا الجدول نستطيع كتابة نموذج الانحدار الموفق كالتالي:

$$\hat{y}_i = 43.513 + 9.929\, x_i \quad , i = 1, 2, \dots, 35$$

ويمكن "مبدئيا" تفسير قيمة معامل الانحدار $\hat{\beta}_1 = 9.929$ بأنه يمثل تأثير معدل ساعات المذاكرة X (المتغير التوضيحي) على درجة الطالب في المقرر 1، Y (المتغير التابع)، بحيث أن كل زيادة بمقدار الوحدة في ساعات المذاكرة، (زيادة ساعة واحدة)، ستؤدي إلى زيادة (نظرا للإشارة الموجبة للمعامل) في درجة الطالب بمقدار 10 درجات تقريبا $(9.929 \cong 10)$، (وذلك بإهمال قيمة الجزء المقطوع $\hat{\beta}_0$). بمعنى أن نمط الزيادة في درجة الطالب في المقرر 1 يتغير بمعامل مقداره 9.929 درجة بحسب نموذج الانحدار الموفق. وبالنسبة للقيم الأخرى في جدول النتائج فإننا سنقدم لها شرحا مفصلا لاحقا.

جدول 7.5: نتائج تقدير معالم نموذج الانحدار الخطي للمتغير "المقرر1" على المتغير "ساعات_الدراسة".

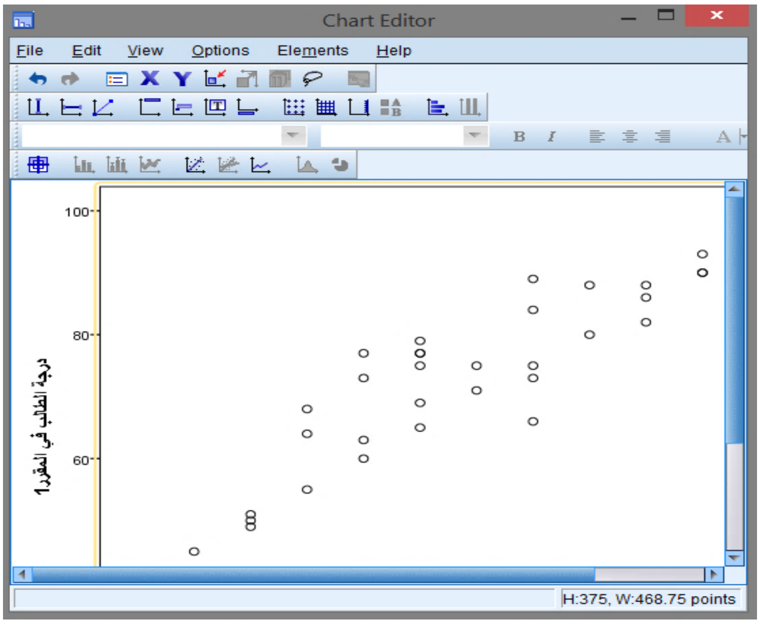
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	43.513	2.660		16.359	.000
	معدل ساعات الدراسة اليومية	9.929	.851	.897	11.669	**.000**

a. Dependent Variable: درجة الطالب في المقرر1

لنقم الآن برسم شكل الانتشار للمتغيرين "المقرر1" و"ساعات_الدراسة" ورسم خط الانحدار الموفق على نفس الرسم[1]. لعمل ذلك، قم في شريط الأوامر العلوي باختيار Graphs>Legacy Dialogs>Scatter/Dot ثم اختر الانتشار البسيط (Simple Scatter) من نافذة الرسم، (شكل (1.5))، ثم اضغط تعريف. وفي نافذة تعريف المتغيرات، (شكل (2.5))، اختر المتغير "المقرر1" وقم بنقله لمربع المحور الصادي (Y-Axis) واختر متغير "ساعات_الدراسة" وانقله لمربع المحور السيني (X-Axis) ثم اضغط موافق.

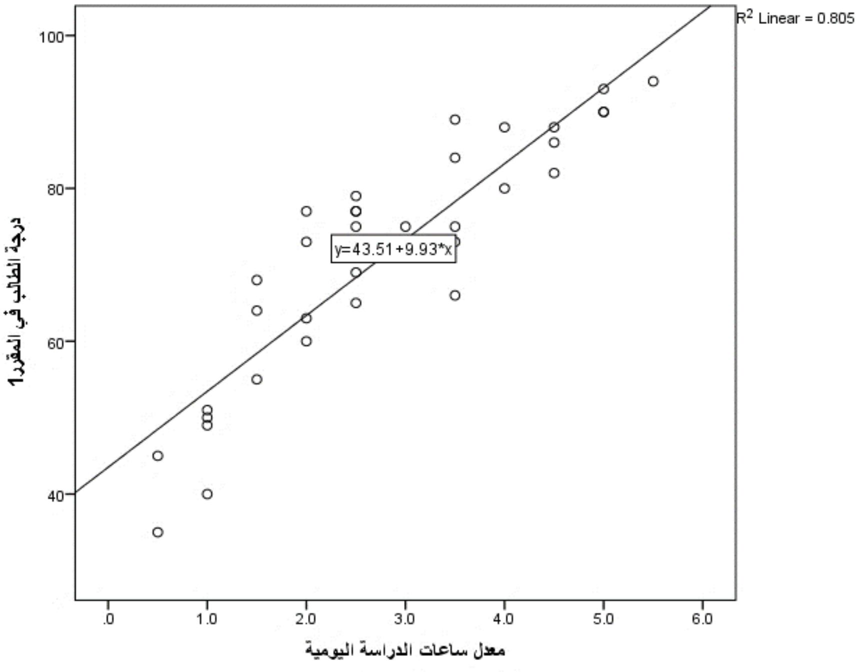
بعد ظهور الرسم في نافذة المخرجات، قم بالنقر المزدوج عليه لتظهر نافذة تحرير الرسم (Chart Editor) كما يظهر في الشكل (17.5)، وفي شريط الأوامر العلوي لتلك النافذة اختر Elements>Fit Line at Total وستلاحظ بعدها ظهور خط الانحدار الموفق في نافذة تحرير الرسم إضافة لظهور نافذة فرعية جديدة بعنوان خصائص (Properties)، والتي تتضمن إمكانية إجراء بعض التعديلات الإضافية، قم بالضغط على إغلاق (Close) في تلك النافذة للعودة لنافذة تحرير الرسم، والتي ستقوم بإغلاقها باستخدام رمز إغلاق النوافذ ذو اللون الأحمر وعلامة ×.



شكل 17.5: نافذة محرر الرسم لشكل الانتشار.

[1] في SPSS خيار الرسم في نافذة تحليل الانحدار يتضمن رسم شكل الانتشار وتوفيق خط الانحدار لقيم المتغير التوضيحي المعيارية (Standardized) وليس القيم الفعلية له، لذلك نلجأ لتمثيل خط الانحدار بهذه الطريقة.

الآن وفي نافذة المخرجات ستلاحظ ظهور خط الانحدار الموفق ومعادلة الانحدار الخطي الموفقة على شكل الانتشار، (شكل (18.5)). ويُلاحظ من الشكل اقتراب الكثير من نقاط البيانات من خط الانحدار مما يدل على انخفاض قيم البواقي وبالتالي وبصورة مبدئية يمكن الحكم على أهمية تأثير المتغير التوضيحي على المتغير التابع في نموذج الانحدار.



شكل 18.5: خط الانحدار الموفق للمتغير "المقرر1" على "ساعات_الدراسة" مع شكل الانتشار لهما.

4.5 تقييم نموذج الانحدار الخطي (Assessment of Linear Regression Model)

بعد تقدير معالم نموذج الانحدار الخطي، نكون بحاجة لتقييم هذا النموذج الموفق بمعنى التأكد من مدى ملائمته وجودته للاستخدام في تفسير العلاقة السببية بين المتغيرات وللتنبؤ ببيانات جديدة. وهذا في العموم هو شأن أي أسلوب إحصائي يعتمد على التقدير. ومن الناحية التطبيقية، فإن هذا التقييم هو في الواقع تقييم لشيئيين أساسيين؛

1. طبيعة ومدى تأثير متغير أو أكثر على المتغير التابع.
2. مدى ملائمة مشاهدات المتغيرين (العينة) للتغير الخطي في العلاقة بينهما.

ولاختبار جودة النموذج سنستعرض الطرق التالية؛

1.4.5 معامل الارتباط المتعدد (Multiple *R*)

رغم أننا لم نتطرق لتحليل الانحدار المتعدد بعد، (وجود أكثر من متغير توضيحي في النموذج)، إلا أنه يمكننا استخدام معامل الارتباط المتعدد، والذي يعتبر معامل الارتباط البسيط حالة خاصة منه عند التعامل مع

متغيرين فقط. والنظرة إلى معامل الارتباط المتعدد بصورة عامة تكون بأنه مقياس للارتباط بين المتغير التابع و"باقي" المتغيرات التوضيحية. وفي حالة نموذج الانحدار البسيط، تكون قيمة معامل الارتباط المتعدد هي نفسها قيمة معامل الارتباط الخطي البسيط (معامل بيرسون) أي أن؛ $R = r_{XY}$. وبالطبع، فإنه كلما زادت قيمة معامل الارتباط المتعدد دل ذلك على قوة العلاقة بين المتغير التابع والمتغير التوضيحي (أو المتغيرات التوضيحية) والذي بدوره يدل على جودة النموذج الموفق. أما بالنسبة للإشارة الجبرية لمعامل الارتباط المتعدد، فإنه يأخذ نفس إشارة معامل الانحدار في نماذج الانحدار البسيط، وتبقى إشارته موجبة في النماذج المتعددة. واحرص، كما ذكرنا سابقا، على ضرورة اختبار معامل الارتباط للتأكد من معنويته.

2.4.5 معامل التحديد R^2 (Coefficient of Determination or *R*-Squared)

معامل التحديد هو مقياس إحصائي يستخدم لقياس جودة توفيق النموذج وهو يمثل نسبة الاختلاف في قيم المتغير التابع التي يفسرها نموذج الانحدار الموفق، أو بمعنى رياضي؛ هو النسبة بين مجموع المربعات للاختلاف المفسر $(\hat{y}_i - \bar{Y})$ ومجموع المربعات للاختلاف الكلي $(y_i - \bar{Y})$، وبالتالي فإنه يأخذ الصيغة التالية؛

$$R^2 = \frac{\sum(\hat{y}_i - \bar{Y})^2}{\sum(y_i - \bar{Y})^2}$$

ورياضيا أيضا فإن معامل التحديد ما هو إلا مربع معامل الارتباط المتعدد، لذلك يمكن كتابة؛

$$R^2 = (R)^2 = \left(\frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}\right)^2 = \frac{S_{XY}^2}{S_{XX}S_{YY}} = \frac{(Cov(X,Y))^2}{Var(X)Var(Y)}$$

والقيمة التي يأخذها معامل التحديد هي نسبة ما يفسره نموذج الانحدار الموفق من العلاقة السببية بين المتغير التابع والمتغير التوضيحي. وكما هو الحال مع معامل الارتباط المتعدد، فإن ارتفاع قيمة معامل التحديد هو مؤشر على صحة أو جودة النموذج.

- **ملاحظات حول معامل التحديد:**

1. ستتراوح قيمة R^2 ما بين 0% و100%، (بعد ضرب قيمته العشرية في 100)، حيث تدل القيمة 0 على أن النموذج لم يفسر أي شيء من الاختلاف في قيم المتغير التابع من خلال نموذج الانحدار، والقيمة 100 تعني أن النموذج الموفق هو "كامل" في توصيف العلاقة السببية، وعادة لا يتم الوصول لهاتين القيمتين من خلال البيانات في الحياة العملية.
2. يمكن مراقبة ما "يعكسه" معامل التحديد بيانيا من خلال رسم شكل الانتشار لقيم المتغير التابع الحقيقية Y ضد قيمه المقدرة $\hat{Y}$.

3. لا يمكن الاعتماد بصورة مطلقة على قيمة معامل التحديد بمفرده للحكم على جودة النموذج لأن القيم المرتفعة للمعامل لا تعني "دائما" أن النموذج جيد، (كما وأن القيم المنخفضة له قد تظهر أحيانا في نماذج ذات معنوية)، لأن قيمته تتعلق بشكل كبير بحجم وطبيعة البيانات وعدد المتغيرات التوضيحية في النموذج، ولذلك يجب أن يتم مراقبة كل نتائج تحليل الانحدار، (ومن ضمنها نتائج تحليل البواقي الذي سنتناوله لاحقا)، قبل تقديم الحكم على النموذج الموفق.
4. يمكن إجراء تعديل على صيغة معامل التحديد لنحصل على **معامل التحديد المعدل** (Adjusted *R*-square)، والذي يُعد أفضل لدى بعض الإحصائيين. وتعطى صيغته بالصورة؛

$$R_{adj}^2 = R^2 - (1 - R^2)\frac{p}{n - p - 1}$$

حيث n هو حجم العينة، و p هو عدد المتغيرات التوضيحية في النموذج.

3.4.5 اختبار معنوية نموذج الانحدار (Testing the Significance of the Regression Model)

هذا الاختبار في الواقع هو اختبار تحليل التباين المعروف باستخدام الإحصاءة فيشر F، وتهدف لاختبار معنوية نموذج الانحدار الموفق ككل، لذلك فإن هذا الاختبار يُعد الأهم ضمن الأساليب المختلفة لتقييم النموذج، حيث يتم اختبار الفرضية الصفرية؛ **(نموذج الانحدار غير معنوي)** $\boldsymbol{H_0}$:.

وإذا تم رفض الفرضية السابقة عند مستوى معنوية $\alpha = 0.05$ ودرجات حرية $\big(p, (n - p - 1)\big)$ فهذا يعني أن نموذج الانحدار الموفق جيد (ذو معنوية) ويعبر عن العلاقة السببية بين المتغيرات. وتعرّف قيمة إحصاءة الاختبار بالصيغة:

$$F_c = \frac{\sum(\hat{y}_i - \bar{Y})^2 / p}{\sum(y_i - \hat{y}_i)^2 / (n - p - 1)}$$

حيث n هي حجم العينة (عدد المشاهدات)، و p هو عدد المتغيرات التوضيحية. أي أن قيمة الإحصاءة F الحسابية هي عبارة عن النسبة بين مجموع المربعات للاختلاف المفسر $(\hat{y}_i - \bar{Y})$ ومجموع المربعات للاختلاف غير المفسر $(y_i - \bar{Y})$ كل مقسوم على درجات الحرية الخاصة به. ويتم عمليا حساب قيمة F من جدول تحليل التباين (ANOVA).

وننوه هنا إلى أن متوسط مجموع مربعات الخطأ هو تقدير تباين حد الخطأ في النموذج؛ $MSE = \hat{\sigma}_\varepsilon^2$. ولحساب قيم مجاميع المربعات الثلاثة الأساسية في جدول تحليل التباين نستخدم الصيغ التالية، والتي تسمى بمجاميع المربعات المصححة (Corrected) والتي تم "ضمن" صيغتها طرح ما يُعرف **بمعامل التصحيح** (Correction Term (*CT*) الذي يساوي $CT = \frac{(\sum Y)^2}{n}$؛

$$\left.\begin{aligned} SSR &= \frac{n\,(Cov(X,Y))^2}{Var(X)} \\ TSS &= n\,Var(Y) \\ SSE &= TSS - SSR \end{aligned}\right\}$$

4.4.5 اختبار معنوية المتغيرات التوضيحية (Testing the Significance of Predictors)

هذا الاختبار يتعلق بمراقبة أهمية تأثير المتغير التوضيحي على المتغير التابع في نموذج الانحدار، (وفي حالة وجود أكثر من متغير توضيحي في النموذج، فإنه يتم تنفيذ هذا الاختبار بعدد هذه المتغيرات)، ويتم اختبار الفرضية الصفرية؛

$$H_0: \beta_j = 0$$

حيث $j = 1, 2, \dots, p$ ، ورفض هذه الفرضية الصفرية عند مستوى معنوية $\alpha = 0.05$ ودرجات حرية $(n - p - 1)$ يدل على أهمية تأثير المتغير التوضيحي ذو الترتيب j في نموذج الانحدار. وتعرف إحصاءة الاختبار، وهي الإحصاءة استيودنت t، بالصيغة التالية:

$$t_c = \frac{|\hat{\beta}_j|}{SE(\hat{\beta}_j)}\,, \qquad j = 1, 2, \dots, p$$

حيث أن؛

$$SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} = \sqrt{\frac{MSE}{(n-1)Var(X)}}$$

وكذلك يمكن إيجاد (تقدير) $(1-\alpha)100\%$ فترة ثقة لمعالم النموذج β_j باستخدام الصيغة؛

$$\beta_j : \hat{\beta}_j \pm t_{\alpha/2}(n - p - 1)\, SE(\hat{\beta}_j)\,, \qquad j = 1, 2, \dots, p$$

ومن فوائد فترة الثقة لمعالم النموذج أنها تزودنا بتقدير للحدود القصوى للقيم التي يمكن أن تأخذها هذه المعالم مما يساعد على تقييم النموذج.

ولنقم الآن بتوفيق نموذج انحدار جديد واستخدام طرق التقييم والاختبار السابقة بصورة عملية في برنامج SPSS، لنفس ملف البيانات "بيانات الطلبة2"، قم في نافذة تحليل الانحدار الخطي، (شكل (15.5))، باختيار المتغير "المقرر1" كمتغير تابع، والمتغير "الأسرة" كمتغير توضيحي، ثم اضغط على خيار الإحصاءات (Statistics) وفي النافذة الفرعية الخاصة به، (شكل (16.5))، قم باختيار عرض التقديرات (Estimates) وفترات الثقة

(Confidence Intervals) و توفيق النموذج (Model Fit) والإحصاءات الوصفية (Descriptives)، ثم اضغط استمرار للعودة للنافذة الأصلية. اضغط بعد ذلك موافق فتظهر النتائج في نافذة المخرجات.

الجدول الأول في تلك النتائج يتضمن الأوساط الحسابية والانحرافات المعيارية وعدد المشاهدات لمتغيري النموذج، والجدول الثاني هو مصفوفة الارتباط، والتي تتضمن معامل ارتباط واحد فقط لأن النموذج يضم متغيرين اثنين فقط، ومن هذا الجدول يمكننا ملاحظة وجود علاقة عكسية قوية ($r_{(\text{المقرر 1})(\text{الأسرة})} = -0.904$) ذات معنوية عالية (P-value = 0). والجدول الثالث يوضح الطريقة المتبعة في إدخال المتغيرات في النموذج.

ومن الجدول الأخير في نافذة المخرجات، (جدول (8.5))، نستطيع كتابة نموذج الانحدار الموفق كالتالي:

$$\hat{y}_i = 98.994 - 4.613\ x_i \quad , i = 1, 2, \dots, 35$$

وهذا يعني أن كل زيادة بمقدار الوحدة في عدد أفراد أسرة الطالب، (زيادة فرد واحد)، ستؤدي إلى انخفاض (نظرا للإشارة السالبة للمعامل) في درجة الطالب بمقدار 5 درجات تقريبا ($4.613 \cong 5$)، وذلك بإهمال قيمة الجزء المقطوع $\hat{\beta}_0$، أي أن الزيادة في عدد أفراد الأسرة يكون له تأثير سلبي على مستوى الطالب الدراسي، وهذا قد يكون بسبب أن اكتظاظ المنزل يؤدي عادة لعدم توفر الجو المناسب للمذاكرة، كما وضحنا سابقا في الجزء الخاص بدراسة الارتباط.

جدول 8.5: نتائج تقدير معالم نموذج الانحدار الخطي للمتغير "المقرر1" على المتغير "الأسرة".

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	98.994	2.549		38.841	.000	93.808	104.179
	عدد أفراد أسرة الطالب	-4.613	.380	-.904	-12.154	**.000**	-5.385	-3.841

a. Dependent Variable: درجة الطالب في المقرر1

نأتي الآن لتقييم جودة نموذج الانحدار البسيط الموفق، فمن نفس الجدول، (جدول (8.5))، نستطيع ملاحظة أن قيمة إحصاءة الاختبار الخاص بمعلمة الانحدار هو $t_c = 38.841$، والقيمة الاحتمالية هي (P-value = 0)، وبالتالي يتم رفض الفرضية الصفرية التي تنص على عدم أهمية تأثير المتغير التوضيحي على المتغير التابع،($H_0: \beta_{(\text{الأسرة})} = 0$)، بمعنى أن عدد أفراد الأسرة له تأثير سلبي هام على مستوى الطالب الدراسي (في المقرر 1).

ومن القيم التي تساعد الباحث في دراسة وتقييم النموذج أيضا هي قيمة الخطأ المعياري لمعامل الانحدار $SE(\hat{\beta}_1) = 0.380$، والتي كلما كانت منخفضة كلما دل ذلك من دقة التقدير[1]. وأما القيمة (Beta =−0.904) فهي قيمة معامل الانحدار المعيارية.

ومن العامودين الأخيرين في الجدول، نستطيع كتابة؛ $-5.385 < \beta_1 < -3.841$ أي أننا يعني أننا واثقون بنسبة 95% بأن تقدير تأثير عدد أفراد الأسرة على مستوى الطالب الدراسي سيتراوح في العموم بين انخفاض من 4 إلى 5 درجات تقريبا في المقرر1.

ومن الجدول الرابع في نافذة المخرجات، (جدول (9.5))، والذي يمثل قيم معامل الارتباط المتعدد ومعامل التحديد ومعامل التحديد المعدل، يمكننا أولا ملاحظة أن قيمة معامل الارتباط المتعدد، والتي تدل على وجود علاقة قوية جدا بين المتغيرين، هي نفسها قيمة معامل بيرسون للارتباط الثنائي، (R = 0.904)، ولكن بدون الإشارة السالبة.

جدول 9.5: قيم معامل التحديد لنموذج انحدار "المقرر1" على "الأسرة".

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	**.904[a]**	**.817**	**.812**	6.769
a. Predictors: (Constant), عدد أفراد أسرة الطالب				

وبالنظر لقيمة معامل التحديد (R Square)، يمكننا القول بأن نموذج الانحدار الموفق يفسر 82% تقريبا من الاختلاف في درجات الطالب في المقرر1 (المتغير التابع) اعتمادا على عدد أفراد الأسرة (المتغير التوضيحي)، وهذه النسبة في التفسير تُعد جيدة جدا، وكذلك الأمر مع قيمة معامل التحديد المعدل الذي يكون له نفس التعليق.

ونأتي الآن للاختبار الأكثر أهمية في تقييم نموذج الانحدار الموفق، أي اختبار النموذج ككل، وهو اختبار تحليل التباين.

جدول 10.5: جدول تحليل التباين لنموذج انحدار "المقرر1" على "الأسرة".

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6767.575	1	6767.575	147.708	**.000[b]**
	Residual	1511.967	33	45.817		
	Total	8279.543	34			
a. Dependent Variable: درجة الطالب في المقرر1						
b. Predictors: (Constant), عدد أفراد أسرة الطالب						

ففي الجدول الخامس في نافذة المخرجات، (جدول (10.5))، يمكننا استنتاج أن نموذج انحدار "المقرر1" على "الأسرة" هو نموذج معنوي، حيث أن $F_c = 147.708$ بقيمة احتمالية P-value = 0 مما يدفعنا لرفض الفرضية الصفرية (نموذج الانحدار غير معنوي) :H_0.

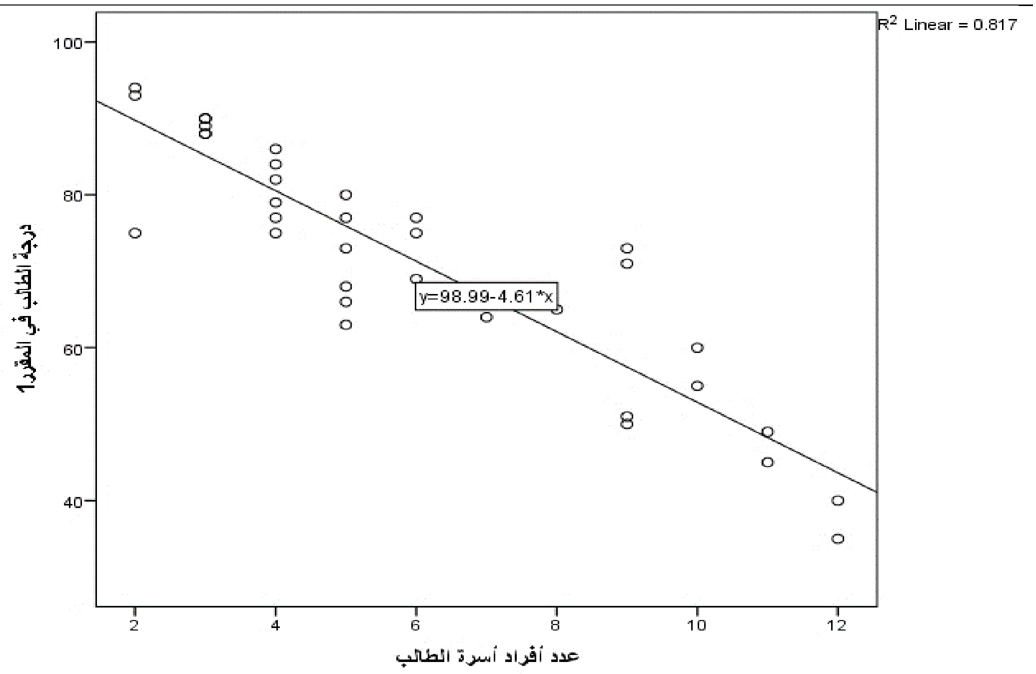
[1] عادة ما تستخدم قيم الخطأ المعياري في المقارنة بين عدة نماذج انحدار لنفس المتغيرات ولكن باستخدام عينات مختلفة أو نماذج يتم فيها تغيير المتغيرات التوضيحية مع "تثبيت" المتغير التابع.

وهكذا، وكمحصلة لنتائج تقييم نموذج الانحدار السابقة، يمكننا استنتاج أن نموذج انحدار درجات الطلبة في المقرر 1 على عدد أفراد الأسرة هو نموذج جيد التوفيق ويمكن استخدامه لدراسة العلاقة بين هذين المتغيرين وللتنبؤ.

فمثلا؛ يمكننا التنبؤ بدرجات طالب ما في المقرر 1 بناء على قيمة افتراضية أو أكثر لعدد أفراد الأسرة، ولنأخذ القيمة $x_0 = 14$ لعدد أفراد أسرة، فعندها يكون: $\hat{y}_0 = 98.994 - 4.613\ (14) = 34.412$، بمعنى أنه إذا أصبح عدد أفراد أسرة أحد الطلبة 14 فردا فإن درجته في المقرر 1 ستنخفض إلى 34 درجة تقريبا.

وختاما لهذا المثال، لنقم برسم خط الانحدار الموفق كما وضحنا أعلاه. قم في شريط الأوامر العلوي باختيار Graphs>Legacy Dialogs>Scatter/Dot ثم اختر الانتشار البسيط، ثم اضغط تعريف. وفي نافذة تعريف المتغيرات اختر المتغير "المقرر1" وقم بنقله لمربع المحور الصادي واختر متغير "الأسرة" وانقله لمربع المحور السيني ثم اضغط موافق.

بعد ظهور الرسم في نافذة المخرجات، قم بالنقر المزدوج عليه لتظهر نافذة تحرير الرسم (Chart Editor)، وفي شريط الأوامر العلوي لتلك النافذة اختر Elements>Fit Line at Total ثم قم بالضغط على إغلاق في تلك النافذة للعودة لنافذة تحرير الرسم، ثم أغلق نافذة التحرير، وسيكون شكل الانتشار للمتغيرين مع خط الانحدار الموفق كما هو في الشكل (19.5)، ويُلاحظ من انتشار النقاط طبيعة العلاقة العكسية بين درجات الطلبة في المقرر 1 وعدد أفراد الأسرة.



شكل 19.5: خط الانحدار الموفق للمتغير "المقرر1" على متغير "الأسرة" مع شكل الانتشار لهما.

ونشير هنا إلى أنه من المفيد أحيانا النظر لقيمة (تقدير) الجزء المقطوع من المحور الصادي، $\hat{\beta}_0$ ، حيث أن القيم المرتفعة لهذا المعامل قد تعني في بعد الأحيان ما يُعرف بنقص توصيف النموذج أي نقص في عدد المتغيرات

التوضيحية اللازمة لتكوين نموذج الانحدار الأفضل. لذلك، وفي كثير من الدراسات عادة ما نتجه لتكوين نماذج انحدار خطي متعددة بهدف مراقبة تأثير عدة متغيرات توضيحية على المتغير التابع، وهذا أقرب للواقع العملي مهما كان مصدر متغيرات النموذج. وهكذا سنتناول التعامل مع نماذج الانحدار الخطي المتعدد في الجزء التالي.

5.5 تحليل الانحدار الخطي المتعدد (Multiple Linear Regression Analysis)

في الكثير من الدراسات التي تتطلب تطبيق أسلوب تحليل الانحدار، نجد أن توصيف العلاقة السببية بين متغيرات الدراسة بمتغيرين فقط عادة ما يكون فيه نقص في هذا التوصيف، بمعنى أن التغير في قيم المتغير التابع سيعتمد في الغالب على أكثر من متغير توضيحي واحد، وهنا تبرز الحاجة لإدراج عدة متغيرات توضيحية في النموذج لتوصيف العلاقة السببية بشكل أوسع. وهكذا يمكن تعريف نموذج الانحدار المتعدد للمتغير التابع على p متغير توضيحي بالصورة التالية؛

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

أو

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i$$

حيث $i = 1, 2, \dots, n$ ، و $j = 1, 2, \dots, p$.

ولنقم باستخدام متغيرات ملف البيانات "بيانات الطلبة2" لتوفيق نموذج انحدار متعدد وتفسير النتائج. ومن الناحية العملية في الدراسات المتضمنة لاستخدام أسلوب تحليل الانحدار المتعدد، يمكننا قبل البدء بتوفيق نموذج الانحدار الخطي المتعدد أن "نستكشف" طبيعة العلاقة بين المتغير التابع وكل متغير من المتغيرات التوضيحية، وذلك عن طريق حساب معامل الارتباط الخطي البسيط لكل زوج منهما.

ولنفرض أننا نريد توفيق نموذج انحدار خطي لدراسة تأثير المتغيرات التوضيحية: درجة الطالب في المقرر3 ("مقرر3")، وعمر الطالب ("العمر")، وعدد أفراد أسرة الطالب ("الأسرة")، ومعدل ساعات دراسة الطالب اليومية ("ساعات_الدراسة") على متغير الاستجابة وهو درجة الطالب في المقرر2 ("مقرر2").

سنقوم أولا بحساب مصفوفة الارتباط للعلاقات الثنائية بين المتغير التابع وكل متغير توضيحي لمراقبة وجود (أو عدم وجود) علاقة خطية بينهما. في شريط الأوامر العلوي قم باختيار Analyze>Correlate>Bivariate وفي نافذة الارتباط البسيط، (الشكل (10.5))، سنقوم باختيار المتغيرات "مقرر2"، "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة" ونقلها إلى مربع المتغيرات (Variables) على اليمين، ثم نضغط موافق.

ستظهر مصفوفة الارتباط المطلوبة في نافذة المخرجات، (جدول (11.5))، وسنقوم باستخدام تلك المصفوفة بتحليل أربعة علاقات فقط، (وهي علاقة المتغير التابع بكل متغير توضيحي)، كما يلي:

جدول 11.5: مصفوفة معاملات الارتباط بين المتغيرات "مقرر2"، "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة".

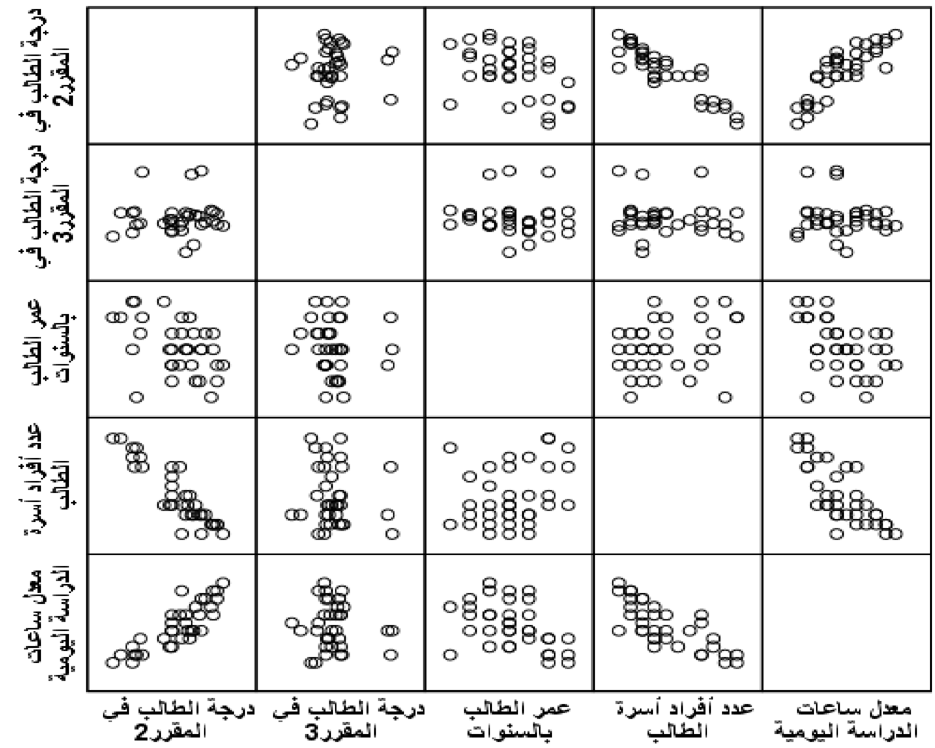
Correlations

		درجة الطالب في المقرر2	درجة الطالب في المقرر3	عمر الطالب بالسنوات	عدد أفراد أسرة الطالب	معدل ساعات الدراسة اليومية
درجة الطالب في المقرر2	Pearson Correlation	1	**.055**	**-.454-****	**-.907-****	**.847****
	Sig. (2-tailed)		**.752**	**.006**	**.000**	**.000**
	N	35	35	35	35	35
درجة الطالب في المقرر3	Pearson Correlation	.055	1	-.082	-.102	-.089
	Sig. (2-tailed)	.752		.639	.561	.610
	N	35	35	35	35	35
عمر الطالب بالسنوات	Pearson Correlation	-.454-**	-.082	1	.292	-.456-**
	Sig. (2-tailed)	.006	.639		.089	.006
	N	35	35	35	35	35
عدد أفراد أسرة الطالب	Pearson Correlation	-.907-**	-.102	.292	1	-.764-**
	Sig. (2-tailed)	.000	.561	.089		.000
	N	35	35	35	35	35
معدل ساعات الدراسة اليومية	Pearson Correlation	.847**	-.089	-.456-**	-.764-**	1
	Sig. (2-tailed)	.000	.610	.006	.000	
	N	35	35	35	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

- نلاحظ وجود علاقة خطية ضعيفة جدا بين درجة الطالب في المقرر2 ودرجته في المقرر3، ($r = 0.055$ و Sig. = 0.752).
- توجد علاقة خطية عكسية معنوية بين درجة الطالب في المقرر2 وعمره، ($r = -0.454$ و Sig. = 0.006).
- وتوجد علاقة خطية عكسية قوية جدا بين درجة الطالب في المقرر2 وعدد أفراد أسرته، حيث ($r = -0.907$ و Sig. = 0).
- كذلك توجد علاقة خطية طردية قوية جدا بين درجة الطالب في المقرر2 معدل ساعات دراسته اليومية، حيث ($r = 0.847$ و Sig. = 0).

ويمكننا إضافة لما سبق، استخدام مصفوفة شكل الانتشار لمراقبة العلاقات السابقة بصورة مرئية كما يظهر في شكل (20.5)، حيث أن العلاقات الأربعة موجودة في الصف الأول من الرسم، ويُلاحظ على الأخص انتشار النقاط بشكل عشوائي غير محدد لدرجة الطالب في المقرر2 ودرجته في المقرر3، ووجود علاقة عكسية "لا بأس بها" بين درجة الطالب في المقرر2 وعمره.



شكل 20.5: مصفوفة شكل الانتشار للمتغيرات "مقرر2"، "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة".

وبالتالي يمكننا مبدئيا القول بأن المتغيرات الثلاثة؛ عمر الطالب وعدد أفراد أسرته ومعدل ساعات دراسته ستكون متغيرات "مؤهلة" لتكوين نموذج انحدار خطي جيد مع متغير درجة الطالب في المقرر2. أما متغير درجة الطالب في المقرر3 فالرأي الأولي يشير بأنه لن يكون ذو أهمية في التأثير على المتغير التابع، إلا أننا، وبدافع تقديم شرح متكامل لتفسير نتائج الانحدار المتعدد، سنكمل توفيق نموذج الانحدار بكل متغيرات الدراسة المقترحة منذ البداية.

لتوفيق نموذج الانحدار المتعدد المطلوب نقوم في نافذة تحليل الانحدار الخطي، (شكل (15.5))، باختيار المتغير "المقرر2" كمتغير تابع، والمتغيرات "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة" كمتغيرات توضيحية، ثم نضغط على خيار الإحصاءات (Statistics) وفي النافذة الفرعية الخاصة به، نختار عرض التقديرات (Estimates) وفترات الثقة (Confidence Intervals) وتوفيق النموذج (Model Fit)، ثم نضغط استمرار للعودة للنافذة الأصلية، ونضغط بعد ذلك موافق فتظهر النتائج في نافذة المخرجات، وسنعرض هنا الجداول الثلاثة المهمة في تلك النافذة تباعا بحسب ترتيب تفسير النتائج المنطقي؛ من جدول (12.5)، نستطيع كتابة نموذج الانحدار المتعدد الموفق كالتالي:

$$\hat{y}_i = 109.369 + 0.007x_{i1} - 1.203x_{i2} - 3.189x_{i3} + 3.220x_{i4} \quad , i = 1, 2, \dots, 35$$

أو بطريقة أخرى؛

$$\widehat{(\text{المقرر}2)}_i = 109.369 + 0.007(\text{المقرر}3)_i - 1.203(\text{العمر})_i - 3.189(\text{الأسرة})_i + 3.220(\text{ساعات_الدراسة})_i$$

ويمكن تفسير قيم معاملات الانحدار بالصورة التالية:

- زيادة درجة واحدة في المقرر3 لن يؤدي لزيادة تُذكر في درجات الطالب في المقرر2، ($\hat{\beta}_1 = 0.007$).
- زيادة سنة واحدة في عمر الطالب يؤدي لانخفض قدره درجة واحدة تقريبا في درجات الطالب في المقرر2، ($\hat{\beta}_2 = -1.203$).
- زيادة فرد واحد في أسرة الطالب يؤدي لانخفض قدره 3 درجات تقريبا في درجات الطالب في المقرر2، ($\hat{\beta}_3 = -3.189$).
- زيادة ساعة واحدة في معدل ساعات المذاكرة اليومية للطالب يؤدي لارتفاع قدره 3 درجات تقريبا في درجات الطالب في المقرر2، ($\hat{\beta}_4 = 3.220$). وذلك بإهمال قيمة الجزء المقطوع $\hat{\beta}_0$ في النموذج.

جدول 12.5: نتائج تقدير معالم نموذج الانحدار الخطي المتعدد للمتغير "المقرر2" على المتغيرات "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة".

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	109.369	17.876		6.118	.000	72.862	145.877
	درجة الطالب في المقرر3	.007	.067	.006	.098	**.923**	-.130	.143
	عمر الطالب بالسنوات	-1.203	.630	-.130	-1.909	**.066**	-2.490	.084
	عدد أفراد أسرة الطالب	-3.189	.481	-.641	-6.633	**.000**	-4.171	-2.207
	معدل ساعات الدراسة اليومية	3.220	1.128	.298	2.854	**.008**	.916	5.525

a. Dependent Variable: درجة الطالب في المقرر2

ولتقييم جودة نموذج الانحدار الموفق، ومن نفس الجدول، (جدول (12.5))، نتجه لقبول الفرضية الصفرية التي تنص على عدم أهمية تأثير المتغيرات "مقرر3" و"العمر" على درجة الطالب في المقرر2، ورفض نفس الفرضية الصفرية بالنسبة للمتغيرين "الأسرة" و"ساعات_الدراسة".

وقبل الوثوب للاستنتاجات النهائية، نكمل تحليل المخرجات، فمن فترات الثقة لمعالم النموذج نستطيع كتابة:

- لمتغير "الأسرة" لدينا: $-4.171 < \beta_3 < -2.207$ أي أننا يعني أننا واثقون بنسبة 95% بأن تقدير تأثير عدد أفراد الأسرة على درجات الطالب في المقرر2 سيتراوح في العموم بين انخفاض من 2 إلى 4 درجات تقريبا.
- لمتغير "ساعات_الدراسة" يكون: $0.916 < \beta_4 < 5.525$ بمعنى أننا يعني أننا واثقون بنسبة 95% بأن تقدير تأثير معدل ساعات الدراسة اليومية على درجات الطالب في المقرر2 سيتراوح في العموم بين ارتفاع من درجة واحدة إلى 6 درجات تقريبا.

أما بالنسبة للمتغيرين التوضيحيين الآخرين فلن يتم التعليق على فترات الثقة الخاصة بمعاملات الانحدار لهما نظرا لعدم أهميتهما في النموذج.

من جدول قيم معامل الارتباط المتعدد ومعامل التحديد ومعامل التحديد المعدل، (جدول (13.5))، يمكن مشاهدة أن قيمة معامل الارتباط المتعدد تدل على وجود علاقة قوية جدا بين المتغير التابع وباقي المتغيرات التوضيحية ككل.

وبالنظر لقيمة معامل التحديد، يمكننا القول بأن نموذج الانحدار الموفق يفسر 89% تقريبا من الاختلاف في درجات الطالب في المقرر2 (المتغير التابع) اعتمادا على المتغيرات التوضيحية في النموذج، وهي نسبة ممتازة للنموذج، وكذلك الأمر مع قيمة معامل التحديد المعدل.

جدول 13.5: قيم معامل التحديد لنموذج انحدار "المقرر1" على المتغيرات "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة".

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	**.945[a]**	**.894**	**.879**	5.284

a. Predictors: (Constant), 3عمر الطالب بالسنوات، عدد ,معدل ساعات الدراسة اليومية، درجة الطالب في المقرر أفراد أسرة الطالب

وأما من نتيجة تحليل التباين لنموذج الانحدار الموفق، (جدول (14.5))، فنخلص إلى أن نموذج انحدار "المقرر2" على باقي المتغيرات التوضيحية هو نموذج معنوي، حيث أن $F_c = 63.015$ بقيمة احتمالية P-value = 0 مما يدفعنا لرفض الفرضية الصفرية (نموذج الانحدار غير معنوي) :H_0.

جدول 14.5: جدول تحليل التباين لنموذج انحدار "المقرر1" على المتغيرات "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة".

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7037.382	4	1759.346	63.015	**.000[b]**
	Residual	837.589	30	27.920		
	Total	7874.971	34			

a. Dependent Variable: 2درجة الطالب في المقرر
b. Predictors: (Constant), معدل ساعات الدراسة اليومية، درجة الطالب في المقرر3، عمر الطالب بالسنوات، عدد أفراد أسرة الطالب

وهكذا، فإنه يمكن القول بأن نموذج انحدار درجات الطلبة في المقرر2 على المتغيرات "مقرر3"، "العمر"، "الأسرة"، و"ساعات_الدراسة" هو نموذج جيد التوفيق في العموم، ولكن يجب أن لا يشتمل على المتغيرين "مقرر3" و"العمر" نظرا لانخفاض تأثيرهما وافتقارهما لعلاقة سببية قوية مع درجات الطلبة في المقرر2. وهذا يعني أن المتغيرات الأكثر أهمية في التأثير على درجة الطالب في المقرر2 هما عدد أفراد الأسرة ومعدل ساعات الدراسة اليومية، ولاحظ أن المتغير التوضيحي "العمر" لم يجتز اختبار المعنوية رغم ارتباطه بالمتغير التابع بشكل معنوي عند حساب معامل الارتباط.

ولاستخدام نموذج الانحدار في التنبؤ، يمكنك التعويض بأي قيم تنبؤيه (للمتغيرات التوضيحية) في النموذج الموفق للحصول على قيمة تقديرية لدرجات الطلبة في المقرر2، ألا أنه من المنطقي إعادة توفيق النموذج أولا باستخدام المتغيرات التي أظهرت تأثيرا معنويا في النتائج السابقة بدلا من استخدام النموذج السابق في التنبؤ. في هذه المرحلة، نقترح على القارئ إعادة توفيق نموذج الانحدار الخطي المتعدد:

1. باستخدام المتغير "مقرر2" كمتغير تابع والمتغيرين "الأسرة"، و"ساعات_الدراسة" كمتغيرات توضيحية وملاحظة التغير في نتائج التحليل.

2. باستخدام متغيرات أخرى من ملف البيانات "بيانات الطلبة2" كمتغيرات تابعة تارة وتوضيحية تارة أخرى ومحاولة الوصول لنماذج جيدة[1].

6.5 تقييم فرضيات النموذج الخطي وتحليل البواقي

(Assessing Model Assumptions and Residual Analysis)

في البند (2.3.5)، تم عرض الفرضيات التي ترتكز عليها طريقة المربعات الصغرى الاعتيادية أو كما يُطلق عليها فرضيات نموذج الانحدار الخطي، وفي هذا البند، سنعرض بعض النقاط الهامة حول كيفية تقييم بعض فرضيات النموذج الخطي والتي سيتعلق كثير منها بتحليل البواقي، التي تمثل ابتعاد قيم المتغير التابع المقدرة عن قيمه الفعلية.

1.6.5 فرضية خطية النموذج (Linearity of the Model Assumption)

حيث أنه لا يمكن رسم شكل الانتشار أو خط الانحدار للنموذج الموفق عند وجود أكثر من متغيرين توضيحيين في النموذج، (بمعنى أننا لن نستطيع تخطي الأبعاد الثلاثة (3D) في تمثيل البياني)، فإننا نلجأ لرسم شكل الانتشار للمتغير التابع مع كل متغير توضيحي على حدة ومراقبة سلوك نقاط الانتشار، فإذا كان نمط انتشار النقاط في "معظم" الأشكال يأخذ السلوك الخطي، (إضافة لوجود معاملات ارتباط خطي معنوية)، فإننا نستطيع القول إن نموذج الانحدار يحقق فرضية الخطية.

[1] في مفهوم تحليل الانحدار الخطي المتعدد توجد عدة طرق لبناء النماذج، ومنها ما يُعرف **باختيار النموذج الأفضل** (Best Model selection) باستخدام مجموعة من المتغيرات، إلا أن المجال لا يسمح هنا بعرضها.

سنستخدم البيانات في جدول (15.5) كتطبيق عملي على تقييم فرضيات النموذج الخطي وتحليل البواقي، وهذه البيانات تضم خمسة متغيرات تمثل نتائج اختبارات تقييم نفسي وسلوكي لعينة مكونة من 25 شخصا في مركز متخصص للتقييم السلوكي والإدراكي، حيث يخضع الشخص لأربعة اختبارات مختلفة، (تقييم1، ...، تقييم4)، ثم يتم منحه التشخيص النفسي النهائي بناء على نتائج التقييمات الأربعة من خلال مؤشرات خاصة، وهذه التقييمات بما فيها التشخيص السلوكي النهائي يتم حسابها على مقياس من 1 إلى 100 درجة.

جدول 15.5: نتائج تقييم نفسي لعينة من 25 شخص.

التشخيص السلوكي	تقييم4	تقييم3	تقييم2	تقييم1	
38	37	50	60	36	1
30	50	49	47	12	2
46	53	53	57	60	3
26	45	43	67	51	4
30	38	45	51	50	5
23	34	45	35	28	6
8	24	30	27	70	7
66	52	66	72	55	8
54	55	56	69	62	9
49	47	55	39	70	10
14	38	40	31	37	11
76	58	63	70	83	12
44	39	46	71	90	13
21	28	48	63	34	14
61	59	59	52	56	15
59	58	52	79	59	16
50	52	50	33	54	17
77	60	57	68	100	18
49	45	58	75	48	19
32	40	45	44	70	20
17	35	41	71	24	21
59	53	64	64	46	22
28	30	43	23	54	23
65	54	65	71	44	24
33	33	47	79	41	25

وارتفاع درجة التشخيص السلوكي تدل على اقتراب الشخص من الوضع السلوكي والنفسي الطبيعي، وانخفاض هذه الدرجة يُعد مؤشرا على وجود اضطرابات سلوكية عند هذا الشخص.

سنقوم أولا بإدخال البيانات من الجدول (15.5) في ملف بيانات جديد في برنامج SPSS باسم "التشخيص السلوكي"، وسيكون الهدف الأساسي من استخدام تحليل الانحدار الخطي لهذه البيانات هو دراسة كيفية تأثير التقييمات الأربع على التشخيص السلوكي النهائي للأشخاص، بمعنى أن متغير "التشخيص_السلوكي" سيمثل المتغير التابع في النموذج، والمتغيرات "تقييم1"، "تقييم2"، "تقييم3"، و"تقييم4" ستكون المتغيرات التوضيحية.

وحيث أننا مهتمون في هذا الجزء بتقييم فرضية خطية نموذج الانحدار، سنقوم أولا بإيجاد مصفوفة معاملات الارتباط الخطي البسيط لمراقبة طبيعة ونوع العلاقة بين المتغير التابع وبقية المتغيرات التوضيحية. في شريط الأوامر العلوي قم باختيار Analyze>Correlate>Bivariate فتظهر نافذة الارتباط البسيط، (الشكل (10.5)). وسنقوم بحساب معاملات الارتباط الخطي البسيط بين كل المتغيرات، فنقوم بنقلها إلى مربع المتغيرات (Variables) على اليمين. وحيث أن كل المتغيرات المختارة هنا هي متغيرات كمية فسنختار حساب معامل بيرسون فقط ثم نضغط موافق فنحصل على مصفوفة الارتباطات البسيطة بين المتغيرات المُختارة كما يظهر في جدول (16.5). وفي هذا الجدول، ستكون مراقبتنا للصف الأخير فقط، أي لمعاملات الارتباط بين المتغير "التشخيص_السلوكي" والمتغيرات التوضيحية الأربعة.

جدول 16.5: مصفوفة الارتباطات البسيطة للمتغيرات "تقييم1"، "تقييم2"، "تقييم3"، "تقييم4"، و"التشخيص_السلوكي".

Correlations

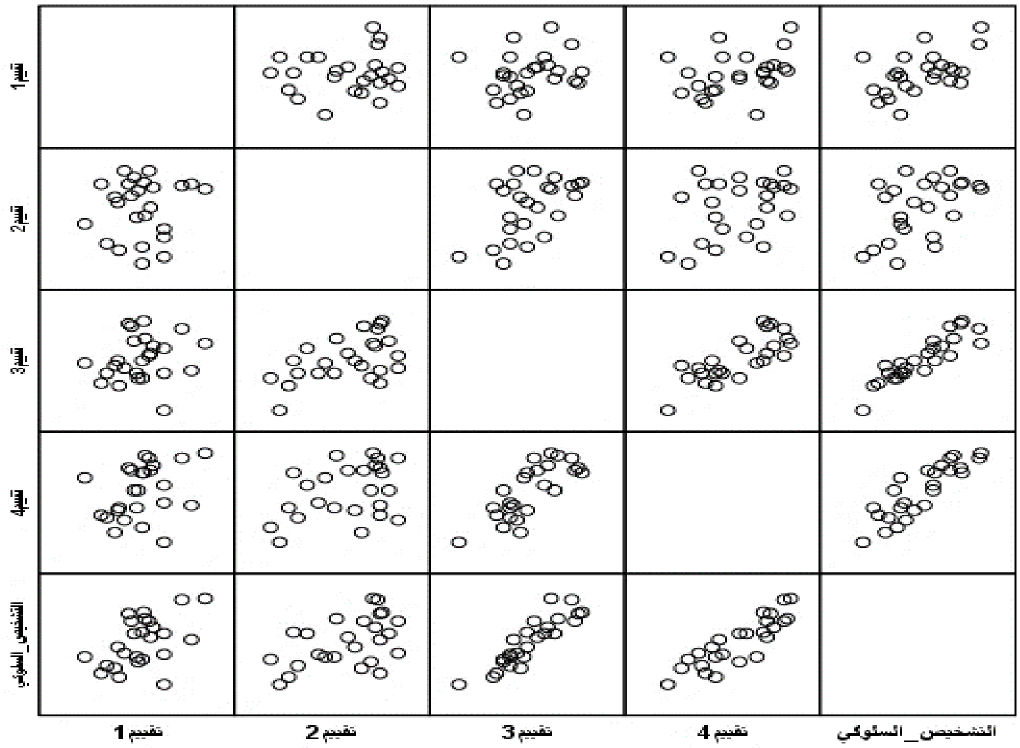
		تقييم1	تقييم2	تقييم3	تقييم4	التشخيص_السلوكي
تقييم1	Pearson Correlation	1	.102	.181	.327	.514**
	Sig. (2-tailed)		.627	.387	.111	.009
	N	25	25	25	25	25
تقييم2	Pearson Correlation	.102	1	.519**	.397*	.497*
	Sig. (2-tailed)	.627		.008	.050	.011
	N	25	25	25	25	25
تقييم3	Pearson Correlation	.181	.519**	1	.782**	.897**
	Sig. (2-tailed)	.387	.008		.000	.000
	N	25	25	25	25	25
تقييم4	Pearson Correlation	.327	.397*	.782**	1	.869**
	Sig. (2-tailed)	.111	.050	.000		.000
	N	25	25	25	25	25
التشخيص_السلوكي	Pearson Correlation	**.514****	**.497***	**.897****	**.869****	1
	Sig. (2-tailed)	**.009**	**.011**	**.000**	**.000**	
	N	25	25	25	25	25

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

ويُلاحظ من اختبارات الفروض الخاصة بتلك المعاملات أن العلاقة بين المتغير التابع وكل المتغيرات التوضيحية هي علاقة خطية ذات معنوية، ($P\text{-value} < 0.05$). وإن كانت علاقة "التشخيص_السلوكي" بالمتغيرين "تقييم3" و"تقييم4" تُعد أقوى بكثير من المتغيرين الآخرين. ولرؤية الصورة بشكل أوضح، سنقوم باستخدام التمثيل البياني لتلك المعاملات، أي أننا سنستخدم مصفوفة شكل الانتشار.

في شريط الأوامر العلوي قم باختيار Graphs>Legacy Dialogs>Scatter/Dot وفي نافذة اختيار شكل الانتشار، قم باختيار شكل مصفوفة الانتشار ثم اضغط تعريف وفي النافذة الفرعية قم باختيار كل المتغيرات ونقلها إلى مربع متغيرات المصفوفة (Matrix Variables) ثم اضغط موافق.

في نافذة المخرجات، سنلاحظ من الصف الأخير في مصفوفة شكل الانتشار، (شكل (21.5))، مدى قوة العلاقة الطردية بين متغير "التشخيص_السلوكي" والمتغيرين "التقييم3" و"التقييم4". إلا أنه في العموم، يمكننا القول أن المتغير التابع في هذا النموذج يرتبط بعلاقات خطية مع المتغيرات التوضيحية، وبالتالي لا توجد مخالفة لفرضية خطية نموذج الانحدار.



شكل 21.5: مصفوفة شكل الانتشار للمتغيرات "تقييم1"، "تقييم2"، "تقييم3"، "تقييم4"، و"التشخيص_السلوكي".

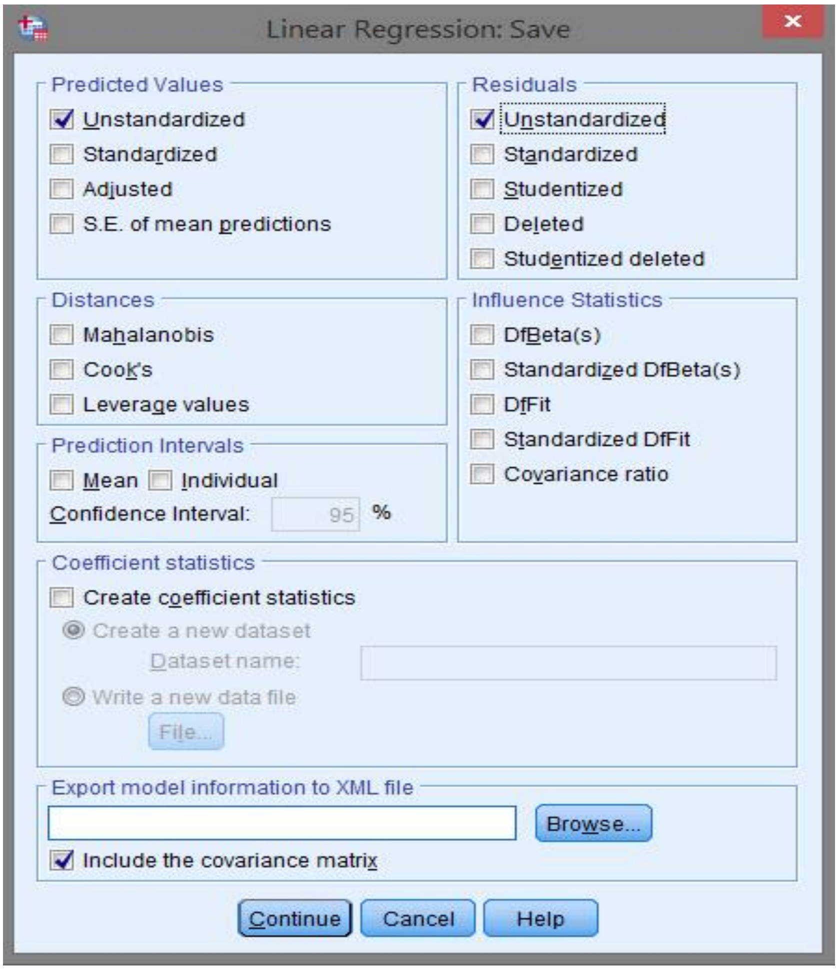
2.6.5 فرضية عدم عشوائية المتغيرات التوضيحية (Non-Random Explanatory Variables)

من ضمن الطرق العملية المستخدمة لتقييم هذه الفرضية هو رسم شكل الانتشار لقيم البواقي ضد قيم كل متغير توضيحي، فإذا أظهر الشكل سلوكا منتظما أو نمطا محددا للنقاط، خطيا كان أو غير خطيا، فهذا يدل على ارتباط هذا المتغير التوضيحي بالخطأ العشوائي مما يدل على عشوائية هذا المتغير وبالتالي مخالفة هذه الفرضية.

ولرسم شكل الانتشار المطلوب، (بين البواقي والمتغيرات التوضيحية)، يجب أولا الحصول على قيم البواقي المقدرة من نموذج الانحدار، $\hat{\varepsilon}$، وهذه القيم يمكن حسابها في SPSS بالطريقة التالية:

في ملف البيانات "التشخيص_السلوكي"، قم بتوفيق نموذج انحدار المتغير "التشخيص_السلوكي" على المتغيرات "التقييم1"، "التقييم2"، "التقييم3"، و"التقييم4"، من خلال اختيار Analyze>Regression>Linear، وفي نافذة تحليل الانحدار الخطي قم بتعيين المتغير التشخيص_السلوكي" كمتغير تابع والمتغيرات "التقييم1"، "التقييم2"، "التقييم3"، و"التقييم4" كمتغيرات توضيحية.

اضغط خيار الحفظ (Save) في هذه النافذة فتظهر نافذة فرعية جديدة كما يظهر في الشكل (22.5)، وهذه النافذة هي مخصصة في معظمها لحفظ القيم المقدرة، وحساب بعض المقاييس من نموذج الانحدار الخطي في ملف البيانات كمتغيرات جديدة.



شكل 22.5: نافذة حفظ القيم المقدرة في نموذج الانحدار الخطي.

قم باختيار حفظ قيم المتغير التابع المقدرة[1] غير المعيارية (Unstandardized) في مربع القيم المقدرة (Predicted Values) واختيار حفظ القيم المقدرة غير المعيارية للبواقي (Unstandardized) في مربع البواقي (Residuals)، كما يظهر في الشكل (22.5)، ثم اضغط استمرار. وفي النافذة الأصلية لتحليل الانحدار اضغط موافق للتنفيذ.

ستظهر بعد ذلك النتائج في نافذة المخرجات، إلا أننا الآن لن نكون مهتمين بها حاليا نظرا لتركيزنا على استخدام قيم البواقي فقط والتي سنجد أنها قد تم إضافتها كمتغير جديد لملف البيانات "التشخيص_السلوكي" باسم "RES_1"، إلى جانب قيم المتغير التابع المقدرة، (باسم "PRE_1")، كما يُلاحظ في الشكل (23.5).

سنقوم الآن برسم شكل الانتشار للبواقي (المتغير "RES_1") مع كل متغير توضحي، أي أننا سنقوم بتنفيذ أربع أشكال انتشار كل على حده.

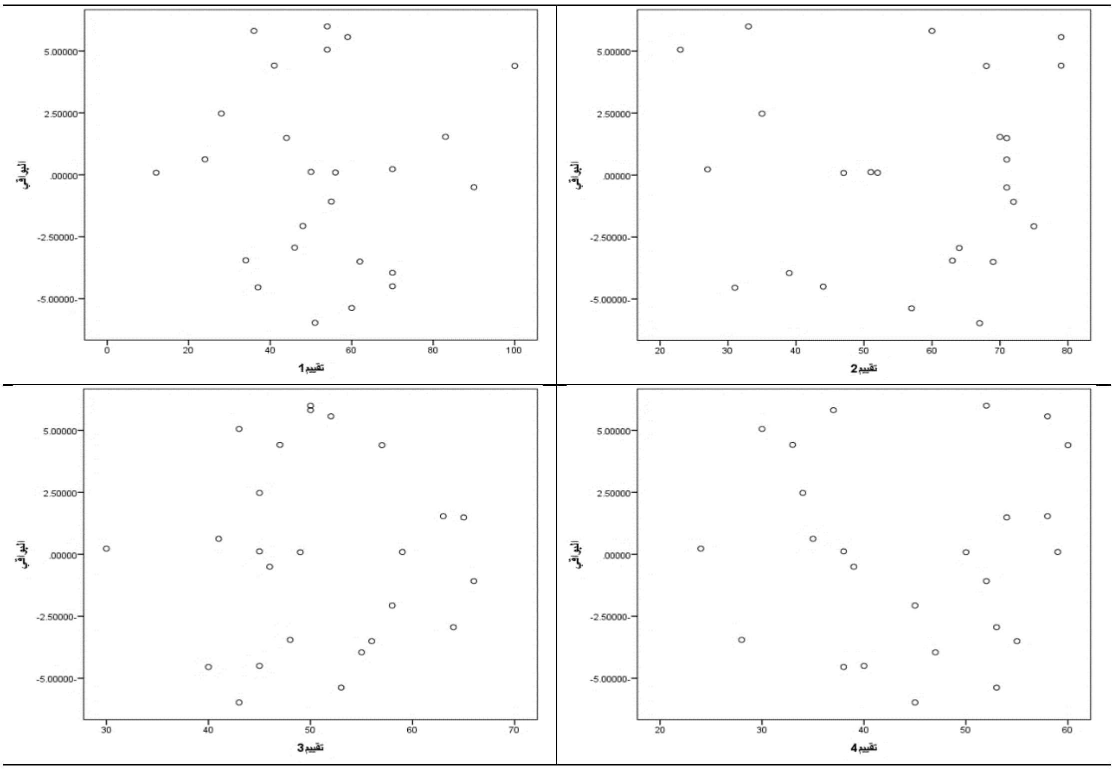
[1] تم اختيار قيم المتغير التابع المقدرة لتوضيح كيفية الحصول عليها في برنامج SPSS، ولن يتم استخدامها في التمثل البياني هنا.

	تقييم1	تقييم2	تقييم3	تقييم4	التشخيص_السلوكي	PRE_1	RES_1
1	36	60	50	37	38	32.18757	5.81243
2	12	47	49	50	30	29.91406	.08594
3	60	57	53	53	46	51.37526	-5.37526-
4	51	67	43	45	26	31.97794	-5.97794-
5	50	51	45	38	30	29.88290	.11710
6	28	35	45	34	23	20.52506	2.47494
7	70	27	30	24	8	7.77087	.22913
8	55	72	66	52	66	67.07927	-1.07927-
9	62	69	56	55	54	57.50383	-3.50383-
10	70	39	55	47	49	52.95644	-3.95644-
11	37	31	40	38	14	18.54266	-4.54266-
12	83	70	63	58	76	74.46389	1.53611

شكل 23.5: ملف البيانات "التشخيص_السلوكي" بعد إضافة قيم البواقي وقيم المتغير التابع المقدرة من نموذج الانحدار.

لتنفيذ شكل الانتشار، سنقوم من شريط الأوامر العلوي باختيار Graphs>Legacy Dialogs>Scatter/Dot وفي نافذة الخيارات، سنبقي على الخيار الافتراضي، وهو شكل الانتشار البسيط، ونضغط على تعريف فتظهر نافذة جديدة لاختيار المتغيرات في شكل الانتشار البسيط، كما في الشكل (2.5)).

في تلك النافذة، سنقوم باختيار متغير البواقي "RES_1" ونقله لمربع المحور الصادي (Y-Axis) واختيار المتغير "تقييم1" ونقله لمربع المحور السيني (X-Axis)، ثم نضغط موافق فنحصل على شكل الانتشار بين المتغيرين. وسنقوم بتكرار هذه الخطوة مع المتغيرات الثلاثة الأخرى؛ "تقييم2"، "تقييم3"، و"تقييم4" فنحصل على الرسومات الأربعة كما يوضح الشكل (24.5).



شكل 24.5: شكل الانتشار للبواقي مع المتغيرات التوضيحية "تقييم1"، "تقييم2"، "تقييم3"، و"تقييم4".

ونلاحظ من شكل انتشار النقاط في الرسومات الأربعة أنها تنتشر بصورة عشوائية ولا تأخذ نمط أو اتجاه معين، وهذا يدل على عدم وجود علاقة، خطية أو غير خطية، بين البواقي وأي متغير توضيحي وبالتالي يدل ذلك على تحقق فرضية عدم عشوائية المتغيرات التوضيحية في نموذج الانحدار الموفق.

3.6.5 فرضية استقلالية المتغيرات التوضيحية (Independence of Errors Assumption)

توجد عدة طرق للتحقق من هذه الفرضية، ولعل أبرزها هي توفيق نماذج انحدار خطي يكون فيها أحد المتغيرات التوضيحية هو المتغير التابع وباقي المتغيرات، (باستثناء المتغير التابع الأصلي)، هي المتغيرات التوضيحية في كل مرة. وبمراقبة نتائج توفيق هذه النماذج الجديدة، يتم تحديد المتغيرات التوضيحية التي تُظهر نتائج "معنوية" كمتغيرات تابعة بأنها متغيرات مرتبطة خطيا بباقي المتغيرات التوضيحية، وهو ما يُعرف بمشكلة **التعدد الخطي** (Multicollinearity) في مفهوم تحليل الانحدار.

للبيانات "التشخيص السلوكي"، سنقوم بالتحقق من استقلالية المتغيرات التوضيحية الأربعة عن بعضها البعض عن طريق توفيق أربعة نماذج انحدار خطي جديدة يأخذ فيها كل متغير توضيحي دور المتغير التابع في كل مرة، ولتنفيذ ذلك في برنامج SPSS، سنقوم في نافذة تحليل الانحدار، التي سنحصل عليها من أمر Analyze>Regression>Linear ، بوضع واحد من المتغيرات التوضيحية؛ "تقييم1"، "تقييم2"، "تقييم3"، و"تقييم4" في مربع المتغير التابع، ووضع باقي المتغيرات، (باستثناء المتغير "التشخيص_السلوكي")، في مربع المتغيرات التوضيحية.

ثم نقوم باختيار الحصول على (Model Fit) فقط من خيار الإحصاءات (Statistics) في نافذة تحليل الانحدار الخطي لكل نموذج من النماذج الأربعة، وقد قمنا بتلخيص نتائج تقييم النماذج الأربعة، (وهي نتائج تحليل التباين ومعاملات التحديد والارتباط المتعدد)، في جدول واحد هو الجدول (17.5).

من هذا الجدول، نستطيع استنتاج ما يلي:

1. المتغيران التوضيحيان "تقييم1" و"تقييم2"، كانت نتائج توفيقهما على باقي المتغيرات التوضيحية غير معنوية؛ (P-value > 0.05)، مصحوبة بمعاملات تحديد متدنية، (0.121 و0.270 على الترتيب)، وبالتالي نستطيع القول بأنهما لا يرتبطان بعلاقات خطية مع باقي المتغيرات التوضيحية، أي أنهما مستقلان خطيا عنها.
2. أما المتغيران التوضيحيان الآخران "تقييم3" و"تقييم4"، فقد أظهرا نتائج معنوية في اعتمادهما خطيا على المتغيرات التوضيحية الأخرى، (P-value = 0)، بمعاملات تحديد مرتفعة، (0.668 و0.647)، وهكذا فإنهما يشكلان مخالفة لفرضية الاستقلال الخطي بين المتغيرات التوضيحية في نموذج الانحدار الأصلي.

جدول 17.5: نتائج تقييم النماذج الأربعة التي تكون فيها المتغيرات التوضيحية "تقييم1"، "تقييم2"، "تقييم3"، و"تقييم4" متغيرات تابعة في كل نموذج.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.348[a]	**.121**	-.004	20.340

a. Predictors: (Constant), تقييم3، تقييم2، تقييم4

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1199.368	3	399.789	.966	**.427**[b]
	Residual	8688.392	21	413.733		
	Total	9887.760	24			

a. **Dependent Variable: تقييم1**
b. Predictors: (Constant), تقييم3، تقييم2، تقييم4

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.519[a]	**.270**	.166	15.797

a. Predictors: (Constant), تقييم3، تقييم1، تقييم4

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1936.458	3	645.486	2.587	**.080**[b]
	Residual	5240.582	21	249.552		
	Total	7177.040	24			

a. **Dependent Variable: تقييم2**
b. Predictors: (Constant), تقييم3، تقييم1، تقييم4

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.818[a]	**.668**	.621	5.451

a. Predictors: (Constant), تقييم2، تقييم1، تقييم4

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1258.108	3	419.369	14.116	**.000**[b]
	Residual	623.892	21	29.709		
	Total	1882.000	24			

a. **Dependent Variable: تقييم3**
b. Predictors: (Constant), تقييم2، تقييم1، تقييم4

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.805[a]	**.647**	.597	6.779

a. Predictors: (Constant), تقييم2، تقييم1، تقييم3

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1770.482	3	590.161	12.843	**.000**[b]
	Residual	964.958	21	45.950		
	Total	2735.440	24			

a. **Dependent Variable: تقييم4**
b. Predictors: (Constant), تقييم2، تقييم1، تقييم3

وننوه هنا إلى إمكانية إجراء اختبار استقلالية المتغيرات التوضيحية خطيا عن بعضها البعض من نافذة تحليل الانحدار، وذلك عند توفيق النموذج المطلوب، باستخدام **معامل تضخم التباين** (Variance Inflation Factor, VIF) المتوفر في خيار الإحصاءات في نافذة تحليل الانحدار، (شكل (16.5))، باسم تشخيص الارتباط الخطي

(Collinearity Diagnostics)، والذي يوفر أيضا مؤشرا جيدا لمراقبة وجود مشكلة التعدد الخطي بين المتغيرات التوضيحية.

4.6.5 فرضية توزع الخطأ بتوزيع طبيعي (Normality of Error Terms)

توجد عدة طرق للتحقق من هذه الفرضية من ضمنها اختبار كولموجروف–سميرنوف واختبار شابيرو–ويلك، واللذان تم التطرق لهما سابقا في الفصل الثالث، إضافة لإمكانية استخدام التمثيل البياني QQ للتحقق من توزع الخطأ طبيعيا.

لدينا في ملف البيانات "التشخيص السلوكي" قيم البواقي لنموذج الانحدار الخطي الموفق، وهكذا سنقوم ببساطة باختيار أمر Analyze>Descriptive Statistics>Explore واختيار خيار الرسم الطبيعي مع الاختبارات (Normality plot with tests) ضمن خيار الرسم (Plots)، ثم الضغط على موافق.

ومن الجدول الخاص باختبارات الطبيعية، (جدول (18.5))، يمكن الاستنتاج من كلا الاختبارين أن البواقي تتبع التوزيع الطبيعي، ($P\text{-value} > 0.05$).

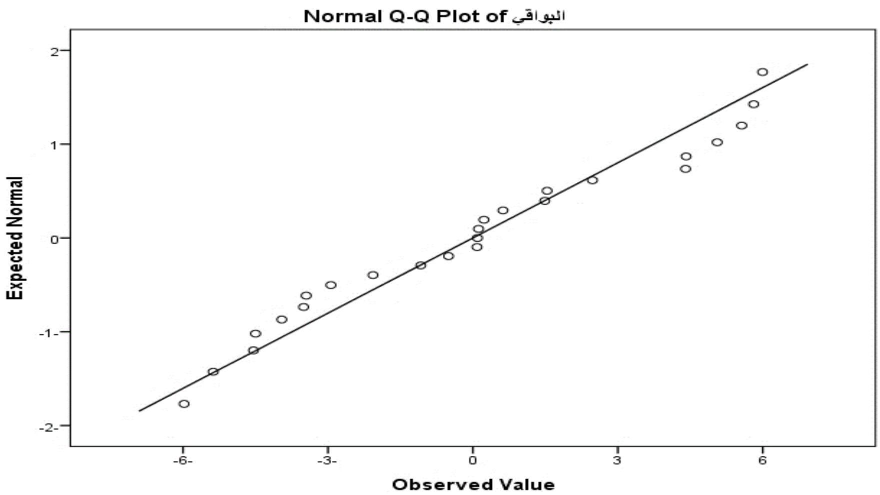
جدول 18.5: اختبارات كولمجروف وشابيرو للبواقي في نموذج انحدار "التشخيص_السلوكي".

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
البواقي	.120	25	**.200***	.943	25	**.174**

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

وإضافة لاختبارات الطبيعية، يمكن استخدام رسم QQ كما يظهر في شكل (25.5)، والذي يؤكد طبيعية توزع البواقي، حيث أن النقاط تقترب من الخط المستقيم بشكل كبير.



شكل 25.5: رسم QQ لبواقي نموذج انحدار "التشخيص_السلوكي".

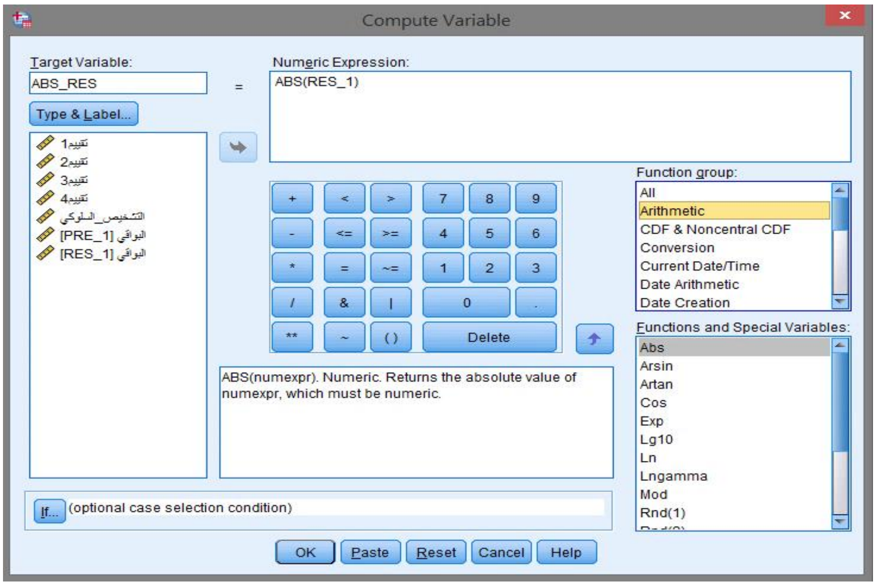
ويمكن الحصول على رسم QQ للبواقي إما باستخدامه من شريط الأدوات العلوي Analyze>Descriptive Statistics>Q-Q Plots، واختيار البواقي "RES_1" في مربع المتغيرات، أو استخدام خيار رسم التوزيع الطبيعي (Normal probability plot) في خيار الرسم (Plots) في نافذة تحليل الانحدار الخطي والذي سيوفر رسم QQ للبواقي المعيارية (Standardized Residuals).

5.6.5 فرضية تجانس تباين قيم حد الخطأ (Homoscedasticity Assumption)

للتحقق من مدى ثبات هذه الفرضية في نموذج الانحدار الخطي، سنستخدم طريقة بسيطة هي طريقة اختبار ارتباط الرتب لسبيرمان، والتي سوف تعتمد على اختبار ارتباط الرتب بين قيم البواقي المطلقة، (أي بأخذ القيمة المطلقة لها)، وقيم المتغيرات التوضيحية في نموذج الانحدار، فإذا وجدت علاقة ذات معنوية بينها دل ذلك على وجود مخالفة للفرضية، أي وجود ما يعرف بمشكلة **عدم التجانس في التباين** (Heteroscedasticity).

في بيانات "التشخيص السلوكي"، سنقوم أولا بتعريف متغير جديد سيكون عبارة عن القيمة المطلقة لقيم البواقي، وسيتم تعريف هذا المتغير كالتالي:

في شريط الأوامر العلوي، سنقوم باختيار Transform>Compute Variable وفي نافذة حساب المتغير سنختار من مجموعة الدوال (Function group) الدوال الرياضية (Arithmetic) ومن تلك الدوال، (في المربع الأسفل منها)، سنختار دالة القيمة المطلقة (Abs)، ونقوم بالنقر المزدوج على الدالة، (أو سحبها بالفأرة إلى مربع الحساب الرقمي (Numeric Expression))، ثم نُدخل متغير البواقي "RES_1" في نطاق الدالة كما هو موضح في الشكل (26.5)، وسنقوم بإعطائه الاسم "ABS_RES" في المتغير المستهدف (Target Variable).



شكل 26.5: نافذة حساب متغير جديد، وهي القيمة المطلقة للبواقي "RES_1" في بيانات "التشخيص_السلوكي".

بعد ذلك نضغط موافق فيتم إدراج المتغير الجديد المطلوب في جدول بيانات "التشخيص السلوكي". الآن سنقوم بتنفيذ اختبار معامل سبيرمان بين قيم البواقي المطلقة وكل متغير توضيحي في النموذج؛

في شريط الأوامر العلوي نختار Analyze>Correlate>Bivariate وفي نافذة الارتباط البسيط نقوم بنقل المتغيرات "تقييم1"، "تقييم2"، "تقييم3"، "تقييم4"، و"ABS_RES" إلى مربع المتغيرات، ونختار حساب معامل سبيرمان فقط من بين المعاملات الثلاثة في (Correlation Coefficients)، ثم نضغط موافق فتظهر مصفوفة الارتباط البسيط كما هو في جدول (19.5).

جدول 19.5: مصفوفة الارتباط البسيط بين المتغيرات "تقييم1"، "تقييم2"، "تقييم3"، "تقييم4"، و"ABS_RES".

Correlations

			تقييم1	تقييم2	تقييم3	تقييم4	ABS_RES
Spearman's rho	تقييم1	Correlation Coefficient	1.000	.030	.230	.441*	.062
		Sig. (2-tailed)		.888	.268	.027	.769
		N	25	25	25	25	25
	تقييم2	Correlation Coefficient	.030	1.000	.474*	.336	-.104
		Sig. (2-tailed)	.888		.017	.101	.622
		N	25	25	25	25	25
	تقييم3	Correlation Coefficient	.230	.474*	1.000	.765**	-.127
		Sig. (2-tailed)	.268	.017		.000	.545
		N	25	25	25	25	25
	تقييم4	Correlation Coefficient	.441*	.336	.765**	1.000	.006
		Sig. (2-tailed)	.027	.101	.000		.977
		N	25	25	25	25	25
	ABS_RES	Correlation Coefficient	.062	-.104	-.127	.006	1.000
		Sig. (2-tailed)	**.769**	**.622**	**.545**	**.977**	
		N	25	25	25	25	25

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

ومن الصف الأخير في مصفوفة الارتباط يتضح أن كل العلاقات الثنائية بين البواقي المطلقة والمتغيرات التوضيحية هي ليست ذات معنوية، (P-value > 0.05)، وبالتالي قبول الفرضية الصفرية ($H_0: \rho_{\hat{\varepsilon}X} = 0$) لكل المتغيرات التوضيحية، مما يدل على تحقق فرضية تجانس تباين حد الخطأ.

الفصل السادس

الاختبارات اللامعلمية وتحليل الموثوقية

(Nonparametric Tests and Reliability Analysis)

في الفصل الرابع، قمنا بالعديد من الاستدلالات الإحصائية حول معالم مثل الوسط الحسابي والتباين، وتم تنفيذ اختبارات لفرضيات تتناول عينة واحدة وعينتين وحتى أكثر، وكذلك تقدير فترات الثقة لتلك المعالم، وهذا يندرج كله تحت ما يُعرف بالتقدير المعلمي (Parametric Estimation) واختبارات الفروض ضمن مفهوم الإحصاء الاستدلالي. والشروط التي يعتمد عليها استخدام هذا التقدير تتمركز في الغالب على كون البيانات (أي العينة أو العينات) المراد اختبارها هي مسحوبة من مجتمعات تتبع توزيعا طبيعيا، إلا أن هذا الشرط أو الافتراض قد لا يتحقق في كثير من الحالات، لذلك نلجأ عندئذ لاستخدام اختبارات أخرى لا تعتمد على التوزيع الاحتمالي للعينة، وهذه الاختبارات تسمى بالاختبارات اللامعلمية (Nonparametric)، وتسمى أحيانا بالاختبارات الحرة من الافتراضات (Assumption-free tests).

ومن ضمن الأنماط التي قد "تُخرج" البيانات من التوزيع الطبيعي كثرة تكرارات القيم المنخفضة أو المرتفعة والذي يسبب التواء التوزيع يمينا أو يسارا، أو كثرة القيم المتطرفة، أو صغر حجم العينة.

ومعظم الاختبارات اللامعلمية ترتكز على مبدأ حساب الرتبة (Ranking) للبيانات، بمعنى إعطاء رتب أو قيم للمشاهدات من الأقل إلى الأكثر، وهكذا فإن القيم الأصغر في البيانات ستأخذ الرتب الأقل، والقيم الأكبر في ستأخذ الرتب الأكبر، ويتم بعد ذلك إجراء الاختبار المطلوب باستخدام الرتب المعطاة بدلا من القيم الفعلية للبيانات. ويمكن تصنيف الاختبارات اللامعلمية إلى ثلاثة أقسام رئيسية، بحسب طبيعة تنظيمها؛

- اختبارات لعينة واحدة.
- اختبارات للعينات المستقلة، (عينتين أو أكثر).
- اختبارات للعينات المرتبطة، (عينتين أو أكثر).

وفي هذا الفصل، سنستعرض بعض أهم الاختبارات اللامعلمية وكيفية تنفيذها في برنامج SPSS، وكذلك قراءة وتفسير النتائج المتعلقة بها.

1.6 الاختبارات اللامعلمية لعينة واحدة (One Sample Nonparametric Tests)

1.1.6 اختبار مربع كاي لجودة التوفيق (Chi-Square Goodness of Fit Test)

عند التعامل مع متغير وصفي، (اسمي كان أو رتبي)، يحتوي على تصنيفات أو تقسيمات معينة (Categories)، قد يكون من المفيد التعرف على النمط الذي تتبعه تكرارات كل تقسيم في هذا المتغير، وما إذا كانت هذه التكرارات متوزعة على التقسيمات بشكل متساوي أو غير متساوي، أو قد نكون مهتمين بالتعرف على التوزيع الاحتمالي الذي قد يتبعه هذا المتغير. في هذه الحالات يمكننا استخدام اختبار جودة التوفيق لمربع كاي لتنفيذ المطلوب.

وتُعرف إحصاءة اختبار مربع كاي لجودة التوفيق بين التكرارات المشاهدة (observed) في المتغير والمتوقعة (expected) له بالصيغة:

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

حيث، o_i هي قيم التكرارات المشاهدة للتقسيم i، e_i هي قيم التكرارات المتوقعة للتقسيم i، و k هو عدد تقسيمات المتغير. وتكون درجات الحرية مساوية لعدد تقسيمات المتغير ناقصاً واحد.

ولتنفيذ اختبار جودة التوفيق في برنامج SPSS لنأخذ المثال التالي؛ في ملف بيانات "فيتامين دال"، لنفرض أننا مهتمون بالتحقق ما إذا كان الأشخاص في العينة:

1. متوزعون على مستويات فيتامين دال الثلاثة؛ (أقل من 30، من 30 إلى 70، وأكثر من 70)، بشكل متقارب (متساوي)، أي اختبار الفرضية الصفرية:

احتمالات توزع الأشخاص في مستويات فيتامين دال **متساوية** :H_0 أو

$H_0: P_{(\text{أقل من } 30)} = P_{(\text{من } 30 \text{ إلى } 70)} = P_{(\text{أكثر من } 70)}$

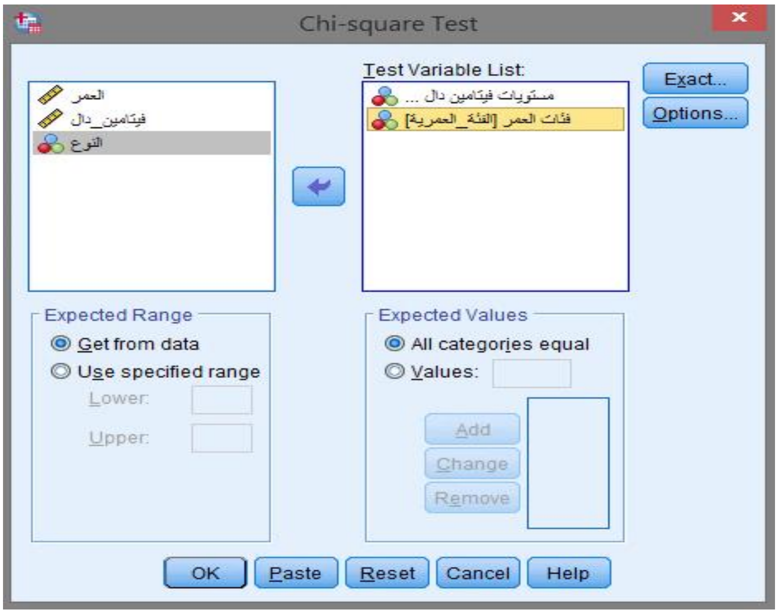
2. متوزعون على الفئات العمرية الثلاثة؛ (أقل من 30، من 30 إلى 50، و 51 فأكثر)، بشكل متقارب، أي اختبار الفرضية الصفرية:

احتمالات توزع الأشخاص في الفئات العمرية **متساوية** :H_0 أو

$H_0: P_{(\text{أقل من } 30)} = P_{(\text{من } 30 \text{ إلى } 50)} = P_{(51 \text{ فأكثر})}$

فنقوم في شريط الأوامر العلوي باختيار Analyze>Nonparametric Tests>Legacy Dialogs>Chi-Square، فتظهر نافذة اختبار مربع كاي كما هو في الشكل (1.6). قم في تلك النافذة باختيار المتغيرين

"مستويات_فيتامين_دال" و"الفئة_العمرية" ونقلهما إلى مربع قائمة متغيرات الاختبار (Test Variable List)، وحيث أننا سنختبر تساوي توزيع الأشخاص في التقسيمات، سنبقي على خيار تساوي التوزيع في التقسيمات (All categories equal) في مربع القيم المتوقعة (Expected values) كما هو، ونضغط موافق للتنفيذ.



شكل 1.6: نافذة اختبار مربع كاي لجودة التوفيق، (اختبار احتمالات متساوية).

في نافذة المخرجات ستلاحظ ظهور النتائج في ثلاثة جداول، (جدول (1.6))، الجدولين الأولين يمثلان توزيع التكرارات المشاهدة والمتوقعة، والفرق بينهما (Residual) للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية" على الترتيب. وأما الجدول الثالث فيضم نتيجة الاختبار. ومن هذا الجدول يمكن رؤية أن:

1. توزيع الأشخاص على مستويات فيتامين دال الثلاثة غير متساوي، حيث أن القيمة الاحتمالية (P-value = 0.005، وبالتالي نرفض H_0).
2. توزيع الأشخاص على الفئات العمرية الثلاثة متساوي، حيث أن (P-value = 0.449، وبالتالي نقبل H_0).

ولاحظ أيضا، من الجدول السابق، ارتفاع قيم الفروقات بين القيم المشاهدة والمتوقعة

جدول 1.6: نتائج اختبار مربع كاي لجودة التوفيق للمتغيرين "مستويات_فيتامين_دال" و"الفئة_العمرية"، (اختبار احتمالات متساوية).

مستويات فيتامين دال

	Observed N	Expected N	Residual
أقل من 30	32	20.0	**12.0**
من 30 إلى 75	14	20.0	**-6.0**
أكثر من 75	14	20.0	**-6.0**
Total	60		

فئات العمر

	Observed N	Expected N	Residual
أقل من 30	16	20.0	-4.0
من 30 إلى 50	20	20.0	0.0
51 فأكثر	24	20.0	4.0
Total	60		

Test Statistics

	مستويات فيتامين دال	فئات العمر
Chi-Square	10.800[a]	1.600[a]
df	2	2
Asymp. Sig.	**.005**	**.449**

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 20.0.

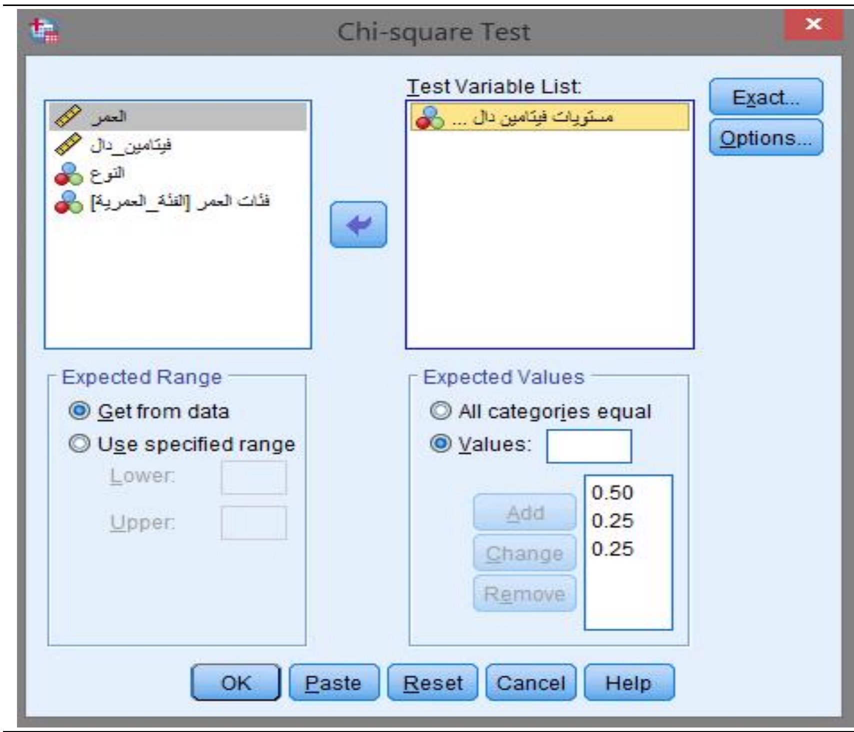
للمتغير "مستويات_فيتامين_دال" والذي يشكل مؤشرا أوليا على أن احتمالات توزع الأشخاص على مستويات الفيتامين الثلاثة لن تكون متساوية.

من جديد، يمكننا استخدام اختبار جودة التوفيق لاختبار ما إذا كانت التكرارات متوزعة على التقسيمات بشكل غير متساوي أي متوزعة باحتمالات غير متساوية، ولنفرض أننا نريد اختبار ما إذا الأشخاص في المثال السابق متوزعون على مستويات فيتامين دال باحتمالات افتراضية معينة، ولتكن الفرضية المراد اختبارها هي:

$$H_0: P_{(\text{أقل من } 30)} = 0.50,\ \ P_{(\text{من } 30 \text{ إلى } 70)} = 0.25,\ \ P_{(\text{أكثر من } 70)} = 0.25$$

أي أننا نريد اختبار فرضية أن الأشخاص متوزعون على المستويات الثلاثة لفيتامين دال باحتمالات أو نسب هي 50%، 25%، و 25% على الترتيب.

لتنفيذ هذا الاختبار، نقوم في شريط الأوامر العلوي باختيار Analyze>Nonparametric Tests>Legacy Dialogs>Chi-Square ، وفي نافذة اختبار مربع كاي نقوم باختيار المتغير "مستويات_فيتامين_دال" ونقله إلى مربع قائمة متغيرات الاختبار، وسنقوم باختيار قيم الاحتمالات (Values) في مربع القيم المتوقعة (Expected values) بحيث نكتب قيمة الاحتمال 0.50 أولا ثم نضغط إضافة (Add) وبعدها نكتب قيمة الاحتمال 0.25 ونضغط إضافة وأخيرا نكتب قيمة الاحتمال 0.25 ونضغط إضافة، فتكون لدينا قيم الاحتمالات كما هو مبين في الشكل (2.6)، ونضغط موافق للتنفيذ.



شكل 2.6: نافذة اختبار مربع كاي لجودة التوفيق، (اختبار احتمالات غير متساوية).

وفي نافذة المخرجات، ستظهر نتائج اختبار مربع كاي، (جدول (2.6))، ويمكن بوضوح ملاحظة أن قيم الفروقات بين القيم المشاهدة والمتوقعة هي ضئيلة، مما ينبئنا بأن توزع الأشخاص على مستويات الفيتامين ستكون على الأغلب كما هو في الفرضية الصفرية، ويدعم ذلك أن القيمة الاحتمالية للاختبار هي (P-value = 0.875)، مما يدفعنا لقبول الفرضية الصفرية القائلة بتوزع الاحتمالات بصورة غير متساوية وهي؛ 0.50، 0.25، و 0.25 بحسب ترتيب مستويات فيتامين دال.

جدول 2.6: نتائج اختبار مربع كاي لجودة التوفيق

للمتغير"مستويات_فيتامين_دال"، (اختبار احتمالات متساوية).

مستويات فيتامين دال

	Observed N	Expected N	Residual
أقل من 30	32	30.0	2.0
من 30 إلى 75	14	15.0	-1.0
أكثر من 75	14	15.0	-1.0
Total	60		

Test Statistics

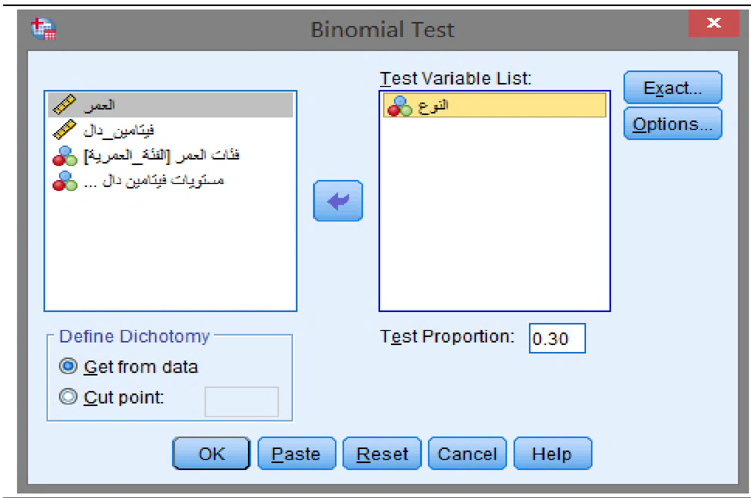
	مستويات فيتامين دال
Chi-Square	.267[a]
df	2
Asymp. Sig.	**.875**

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 15.0.

2.1.6 اختبار ذي الحدين (Binomial Test)

من الاختبارات اللامعلمية المتوفرة في برنامج SPSS هو اختبار ذي الحدين والذي يتم فيه اختبار ما إذا كان متغير وصفي ثنائي، (أي له تقسيمين أو مستويين فقط)، يتبع توزيع ذي الحدين باحتمال معين.

لنفس البيانات؛ "فيتامين دال"، سنقوم باختبار ما إذا كان متغير "النوع" يتبع توزيع ذي الحدين باحتمال 0.30 للذكور، (احتمال النجاح)، واحتمال 0.70 للإناث، (احتمال الفشل). عندها نقوم في شريط الأوامر العلوي باختيار الاختبار Analyze>Nonparametric Tests>Legacy Dialogs>Binomial، فتظهر نافذة الاختبار كما هو موضح في الشكل (3.6).



شكل 3.6: نافذة اختبار ذي الحدين لمتغير "النوع" باحتمال نجاح 0.30.

قم باختيار متغير "النوع" ونقله إلى مربع قائمة متغيرات الاختبار (Test Variable List)، ثم قم بتغيير احتمال النجاح الافتراضي (Test

Proportion) من 0.50 إلى 0.30 كما هو مبين في الشكل. ثم اضغط موافق للتنفيذ. وستحصل على نتيجة الاختبار في نافذة المخرجات كما يظهر في جدول (3.6). ومن هذا الجدول يمكن ملاحظة أن (P-value = 0) وبالتالي نرفض الفرضية الصفرية القائلة بأن متغير "النوع" يتبع توزيع ذي الحدين باحتمال نجاح يساوي 0.30، أو (H_0: P = 0.30).

جدول 3.6: نتيجة اختبار ذي الحدين لمتغير "النوع" باحتمال نجاح 0.30، (للبيانات "فيتامين دال").

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
النوع	Group 1	ذكر	31	.5	.3	**.000**
	Group 2	أنثى	29	.5		
	Total		60	1.0		

وبالطبع هذا لا يعني أن متغير النوع لا يتبع توزيع ذي الحدين، ولكن هذه النتيجة تعني أن احتمال النجاح الافتراضي (0.30) لهذا المتغير هو غير متوافق مع المشاهدات.

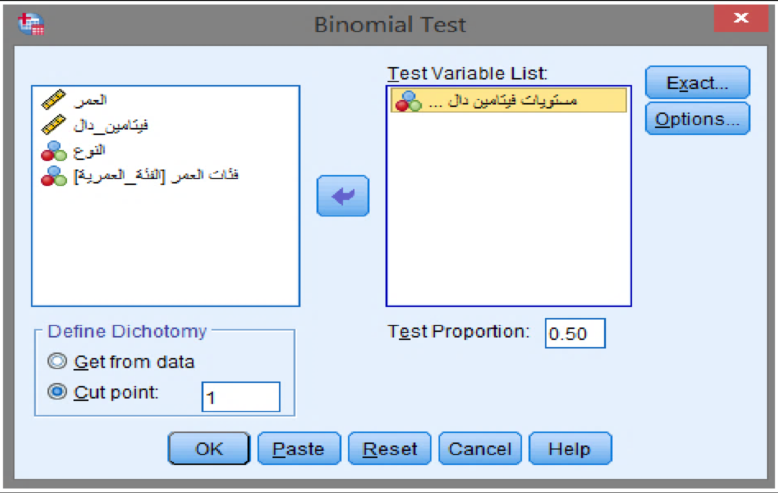
ملاحظة:

قم بإعادة تنفيذ اختبار ذي الحدين السابق مستخدما احتمال نجاح يساوي 0.50 وستجد أن نتيجة الاختبار ستتغير إلى قبول الفرضية (H_0: P = 0.50)، لأن القيمة الاحتمالية للاختبار ستكون P-value = 0.897.

في كثير من الأحيان قد نتعامل مع متغيرات وصفية غير ثنائية، أي أن المشاهدات فيها تأخذ أكثر من مستويين، أو أننا قد نتعامل مع متغير كمي ونرغب في اختبار ما إذا كان يتبع توزيع ذي الحدين إذا ما تم تقسيم قيم المشاهدات إلى قسمين بناء على قيمة محددة، عندئذ يمكننا استخدام اختبار ذي الحدين بالصورة التالية:

1. في بيانات "فيتامين دال"، سنقوم باختبار توزع المتغير الوصفي الرتبي "مستويات_فيتامين_دال"، والذي يأخذ ثلاثة مستويات، بتوزيع ذي الحدين عن طريق اعتبار أن الأشخاص الذين مستويات الفيتامين لديهم أقل من 30 يتبعون التقسيم الأول الجديد، والذين لديهم مستويات فيتامين 30 فأكثر، (وتشمل المستويين الأصليين من 30 إلى 75، وأكثر من 75)، يتبعون التقسيم الثاني الجديد.

وبالتالي سنقوم في نافذة اختبار ذي الحدين باختيار المتغير "مستويات_فيتامين_دال" ونقله إلى مربع قائمة متغيرات الاختبار ولإعادة تقسيم المتغير ذو الثلاثة مستويات إلى المستويين المذكورين أعلاه، سنقوم باختيار نقطة الفصل (Cut point) في مربع تعريف المستويين الثنائيين (Define Dichotomy)، ثم ندخل القيمة 1 (والتي ترمز للأشخاص الذين لديهم مستويات فيتامين أقل من 30) كقيمة للفصل بين المستويات. وسنبقي قيمة الاحتمال الاختبارية (Test Proportion) عند القيمة 0.50، كما هو موضح في الشكل (4.6)، ثم نضغط موافق.



شكل 4.6: نافذة اختبار ذي الحدين لمتغير "مستويات_فيتامين_دال" باحتمال نجاح 0.50.

وفي نافذة المخرجات، ستظهر نتيجة اختبار ذي الحدين، (جدول (4.6))، ونلاحظ أن (P-value = 0.699) وبالتالي نقبل الفرضية الصفرية القائلة بأن متغير "مستويات_فيتامين_دال" يتبع توزيع ذي الحدين باحتمال نجاح يساوي 0.50، أو (H_0: P = 0.50). وهكذا يمكن القول بأن احتمال أن يكون مستوى فيتامين دال لأي شخص هو 30 فأقل يساوي 0.50.

جدول 4.6: نتيجة اختبار ذي الحدين لمتغير "مستويات_فيتامين_دال" باحتمال نجاح 0.50، (للبيانات "فيتامين دال").

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
مستويات فيتامين دال	Group 1	<= 1	32	.53	.50	**.699**
	Group 2	> 1	28	.47		
	Total		60	1.00		

2. من جديد في بيانات "فيتامين دال"، سنقوم باختبار توزع المتغير الكمي "العمر"، بتوزيع ذي الحدين عن طريق اعتبار أن الأشخاص الذين أعمارهم هي 45 سنة فأقل يتبعون التقسيم الأول الجديد، والذين أعمارهم أكبر من 45 سنة يتبعون التقسيم الثاني الجديد.

سنقوم الآن في نافذة اختبار ذي الحدين باختيار المتغير "العمر" ونقله إلى مربع قائمة متغيرات الاختبار ولتقسيم المتغير الكمي إلى مستويين سنقوم باختيار نقطة الفصل (Cut point) في مربع تعريف المستويين الثنائيين (Define Dichotomy)، ثم ندخل القيمة 45 كقيمة للفصل بين الأعمار. وسنبقي قيمة الاحتمال الاختبارية (Test Proportion) عند القيمة 0.50، ثم نضغط موافق.

في نافذة المخرجات، تُظهر نتيجة اختبار ذي الحدين، (جدول (5.6))، أن (P-value = 0.897) وبالتالي نقبل الفرضية الصفرية القائلة بأن متغير العمر يتبع توزيع ذي الحدين باحتمال نجاح يساوي 0.50. وهكذا يمكن القول بأن احتمال أن يكون عمر أي شخص هو 45 فأقل يساوي 0.50.

جدول 5.6: نتيجة اختبار ذي الحدين لمتغير "العمر" باحتمال نجاح 0.50، (للبيانات "فيتامين دال").

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
العمر	Group 1	<= 45	29	.48	.50	**.897**
	Group 2	> 45	31	.52		
	Total		60	1.00		

ملاحظة:

قم بإعادة تنفيذ اختبار ذي الحدين السابق لمتغير العمر مستخدما القيمة 60 سنة كقيمة للفصل بين المستويين، وستجد أن نتيجة الاختبار ستتغير إلى رفض الفرضية الصفرية (H_0: P = 0.50)، لأن القيمة الاحتمالية للاختبار ستكون P-value = 0.006.

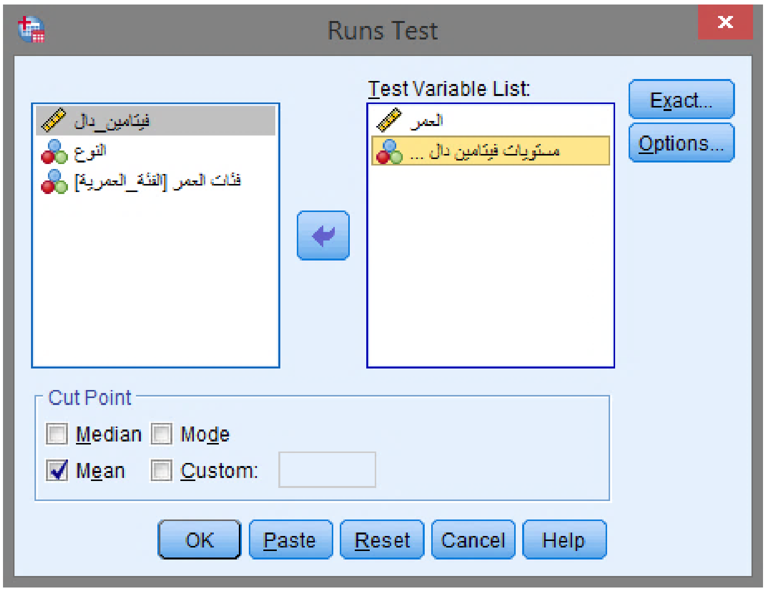
3.1.6 اختبار الدورات (Runs Test)

اختبار الدورات أو اختبار والد–ولفويتز للدورات (Wald-Wolfowitz runs test) هو اختبار لا معلمي يهدف للتحقق من مدى عشوائية أو استقلالية المشاهدات المتتالية ضمن متغير ما. بمعنى آخر، اختبار الدورات قد يساعد الباحث في التحقق من عشوائية المعاينة التي تمت من المجتمع. وتكون الفرضية الصفرية التي يتم اختبارها هي: كل مشاهدة في متتالية العينة قد تم سحبها بصورة مستقلة من نفس المجتمع.

ويعتمد تطبيق اختبار الدورات من الناحية النظرية على إعطاء إشارة موجبة (+) لقيم المشاهدات الأكبر من قيمة معينة، (والتي يمكن أن تكون قيمة الوسط الحسابي أو الوسيط أو المنوال أو أي قيمة أخرى يحددها المستخدم)، وإعطاء الإشارة السالبة (–) للقيم الأقل من تلك القيمة المحددة. ثم يتم حساب عدد الدورات عن طريق حساب عدد التغيرات في الإشارات الموجبة والسالبة، ويتم بعد ذلك مقارنة عدد الدورات المحسوبة من المشاهدات بعدد الدورات المتوقع والذي يتبع، تحت الفرضية الصفرية، توزيعا طبيعيا تقريبا.

ولتطبيق اختبار الدورات في برنامج SPSS، نقوم بفتح ملف البيانات "فيتامين دال" ثم نختار من شريط الأوامر العلوي Analyze>Nonparametric Tests>Legacy Dialogs>Runs فتظهر نافذة اختبار الدورات، قم في تلك النافذة باختيار المتغيرين "العمر" و"مستويات_فيتامين_دال" على سبيل المثال وانقلها إلى مربع قائمة متغيرات الاختبار، ثم اختر الوسط الحسابي (Mean) فقط[1] من مربع قيمة القطع (Cut Point) كما يظهر في الشكل (5.6)، ثم ضغط موافق.

[1] يوفر برنامج SPSS ثلاثة خيارات أساسية لقيم القطع هي الوسط الحسابي، الوسيط، والمنوال، إضافة للقيمة التي يحددها المستخدم (Custom)، وننوه هنا إلى أن استخدام قيم الوسيط و/أو المنوال قد لا يكون ملائما في بعض الحالات، خاصة في المتغيرات التي تأخذ قيم ثنائية.



شكل 5.6: نافذة اختبار الدورات للمتغيرين "العمر" و"مستويات_فيتامين_دال".

ستظهر بعد ذلك النتائج للمتغيرين في نافذة المخرجات، (جدول (6.6))، ونتجه منها لرفض الفرضية الصفرية القائلة بأن المشاهدات المتتالية في متغير العمر قد تم سحبها بصورة مستقلة من المجتمع، بمعنى أن نتيجة اختبار الدورات يفترض أن أعمار الأشخاص المتتالية في العينة لم يتم جمعها بصورة عشوائية.

جدول 6.6: نتائج اختبار الدورات للمتغيرين "العمر" و"مستويات_فيتامين_دال".

Runs Test

	العمر	مستويات فيتامين دال
Test Value[a]	47.57	1.70
Cases < Test Value	32	32
Cases >= Test Value	28	28
Total Cases	60	60
Number of Runs	23	32
Z	-2.058	.296
Asymp. Sig. (2-tailed)	**.040**	**.767**

a. Mean

وعلى النقيض، نميل لقبول الفرضية القائلة بأن المشاهدات (التقسيمات) المتتالية في متغير مستويات فيتامين دال هي مستقلة، (عشوائية في تسلسلها في العينة).

ملاحظة:

قم بإعادة تنفيذ اختبار الدورات السابق لمتغيرات أخرى سواء في نفس ملف البيانات أو ملف آخر مستخدما قيم قطع متنوعة؛ قيمة الوسيط، المنوال، وقيم أخرى، ولاحظ الفروقات بين النتائج.

4.1.6 اختبار كولمجروف–سميرنوف لعينة واحدة (One-Sample Kolmogorov-Smirnov Test)

تناولنا فيما سبق استخدام اختبار كولمجروف–سميرنوف لاختبار توزع المتغير (أو العينة) بتوزيع طبيعي، والأمر هنا مشابه حيث أننا سنقوم باستخدام هذا الاختبار للتحقق من تبعية المتغير الكمي للتوزيع الطبيعي أو التوزيع المنتظم (Uniform)، أو توزيع بواسون (Poisson)، أو التوزيع الأسي (Exponential)، وهي التوزيعات التي يوفرها اختبار كولمجروف–سميرنوف لعينة واحدة في برنامج SPSS.

ويعتمد اختبار كولمجروف–سميرنوف على مقارنة التوزيع الاحتمالي (التراكمي) الفعلي للمتغير بالتوزيع الاحتمالي المطلوب (المتوقع)، وكلما انخفضت قيمة احصاءة الاختبار (اقتربت من الصفر) كلما دل ذلك على اقتراب التوزيع الاحتمالي الفعلي للمتغير من التوزيع الاحتمالي المطلوب مقارنته به. وتكون الفرضية الصفرية للاختبار أن المتغير يتبع التوزيع الاحتمالي المفترض، (الطبيعي، المنتظم، ...).

سنستخدم ملف البيانات "بيانات الطلبة" للتحقق من التوزيع الاحتمالي لبعض المتغيرات، وبالتالي قم في شريط الأدوات العلوي باختيار Analyze>Nonparametric Tests>Legacy Dialogs>1-Sample K-S، فتظهر نافذة الاختبار، كما يوضح الشكل (6.6). قم في تلك النافذة باختيار المتغيرات؛ "المقرر1"، "المقرر3"، و"الفصل" على سبيل المثال، ثم انقلها إلى مربع قائمة متغيرات الاختبار. ولنفرض أننا نريد اختبار ما إذا كان كل متغير من هذه المتغيرات يتبع التوزيع الطبيعي أو التوزيع المنتظم، لذلك سنقوم باختيار طبيعي (Normal) ومنتظم (Uniform) من مربع التوزيع الاحتمالي للاختبار (Test Distribution) كما يوضح الشكل، ثم نضغط موافق للتنفيذ.



شكل 6.6: نافذة اختبار كولمجروف–سميرنوف لعينة واحدة، (للبيانات "معدلات الطلبة").

في نافذة المخرجات، ستكون النتائج موضحة في جدولين؛ الأول خاص بنتائج المقارنة بالتوزيع الطبيعي، (لكل متغير من المتغيرات الثلاثة المختارة)، والثاني خاص بنتائج المقارنة بالتوزيع المنتظم. ونلاحظ من الجدولين، (اللذين قد تم عرضهما في جدول واحد هو جدول (7.6))، ما يلي:

1. المتغير درجات الطلبة في المقرر 1 يتوزع بتوزيع طبيعي تقريبا، (P-value = 0.200 في اختبار التوزيع الطبيعي)، ولا يتوزع بالتوزيع المنتظم، (P-value = 0.029 في اختبار التوزيع المنتظم).

2. المتغير درجات الطلبة في المقرر 3 لا يتوزع بتوزيع طبيعي، (P-value = 0 في اختبار التوزيع الطبيعي)، ولا يتوزع بالتوزيع المنتظم، (P-value = 0 في اختبار التوزيع المنتظم).

3. المتغير ترتيب الفصل الدراسي للطالب يتوزع بتوزيع طبيعي تقريبا وإن كانت القيمة الاحتمالية للاختبار قريبة جدا من القيمة 0.05، (P-value = 0.060 في اختبار التوزيع الطبيعي)، ويتوزع بالتوزيع المنتظم، (P-value = 0.751 في اختبار التوزيع المنتظم)، لذلك نتجه لاعتباره أقرب للتوزيع المنتظم.

جدول 7.6: نتائج اختبار كولمجروف–سميرنوف لعينة واحدة للمقارنة بالتوزيع الطبيعي والتوزيع المنتظم، للمتغيرات "المقرر1"، "المقرر3"، و"الفصل"، (للبيانات "معدلات الطلبة").

One-Sample Kolmogorov-Smirnov Test

		درجة الطالب في المقرر1	درجة الطالب في المقرر3	ترتيب الفصل الدراسي للطالب
N		35	35	35
Normal Parameters[a,b]	Mean	71.31	55.34	5.06
	Std. Deviation	15.605	14.289	1.846
Most Extreme Differences	Absolute	.114	.235	.145
	Positive	.075	.235	.145
	Negative	-.114	-.129	-.124
Test Statistic		.114	.235	.145
Asymp. Sig. (2-tailed)		**.200**[c,d]	**.000**[c]	**.060**[c]

a. Test distribution is Normal.
b. Calculated from data.
c. Lilliefors Significance Correction.
d. This is a lower bound of the true significance.

One-Sample Kolmogorov-Smirnov Test 2

		درجة الطالب في المقرر1	درجة الطالب في المقرر3	ترتيب الفصل الدراسي للطالب
N		35	35	35
Uniform Parameters[a,b]	Minimum	35	27	2
	Maximum	94	95	8
Most Extreme Differences	Absolute	.246	.400	.114
	Positive	.029	.400	.095
	Negative	-.246	-.150	-.114
Kolmogorov-Smirnov Z		1.455	2.364	.676
Asymp. Sig. (2-tailed)		**.029**	**.000**	**.751**

a. Test distribution is Uniform.
b. Calculated from data.

2.6 الاختبارات اللامعلمية لعينتين مستقلتين

(Nonparametric Tests for Two Independent Samples)

يوفر برنامج SPSS الاختبارات التالية في منطقة الاختبارات اللامعلمية لعينتين مستقلتين:

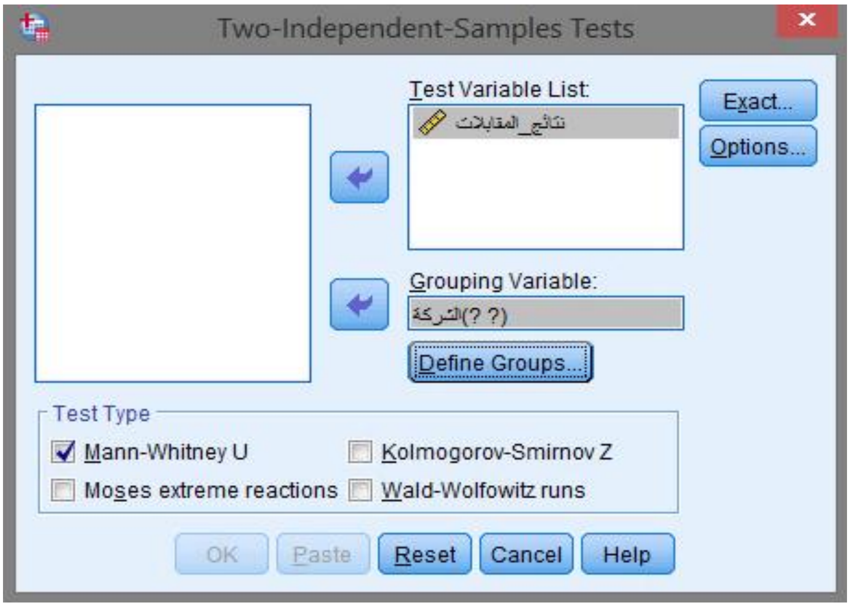
1. اختبارات **مان–ويتني U وويلكوكسون لمجموع الرتب W** (**Mann-Whitney U and Wilcoxon rank sum W** Tests).
2. اختبار **كولمجروف–سميرنوف Z** (**Kolmogorov-Smirnov Z** Test).
3. اختبار **والد–ولفويتز للدورات** (**Wald-Wolfowitz Runs** Test).
4. اختبار **موسز للردود المتطرفة** (**Moses Extreme Reactions** Test).

يُعد اختبار مان–ويتني "النسخة" اللامعلمية من اختبار t لتساوي متوسطي المجتمع، إذا يقوم باختبار الفرضية القائلة بتساوي توزيعي عينتين مستقلتين، أو بمعنى آخر، يقوم الاختبار بالتحقق من أن العينتين مسحوبتين من مجتمعين متماثلين. ويُعتبر اختبار ويلكوكسون لمجموع الرتب مكافئ لاختبار مان–ويتني لاختبار عينتين. أي أن كل من الاختبارين يقوم باختبار الفرضية الصفرية: **العينتين لهما نفس التوزيع :H_0**.

وأما اختباري كولمجروف–سميرنوف ووالد–ولفويتز للدورات فقد تم التطرق لهما سابقا، ويقوما باختبار نفس الفرضية الصفرية أعلاه عن طريق التعامل مع كل عينة أو مجموعة على حده ثم مقارنة النتائج.

بالنسبة لاختبار موسز للردود المتطرفة، فيقوم باعتبار أن العينة أو المجموعة الأولى من المشاهدات هي مجموعة التحكم (Control group) واعتبار العينة الثانية هي مجموعة التجربة (Experimental group)، ثم يقوم بدمج العينتين وإعطاء رتب لكل المشاهدات فيهما وتنفيذ الاختبار.

ولتطبيق هذه الاختبارات، قم بفتح ملف البيانات "المقابلات الشخصية"، والذي سيكون المثال التطبيقي لنا في برنامج SPSS. ثم قم في شريط الأوامر العلوي باختيار Analyze>Nonparametric Tests>Legacy Dialogs>2 Independent Samples، فتظهر نافذة الاختبار كما هو في الشكل (7.6). في تلك النافذة، قم بنقل المتغير "نتائج_المقابلات" إلى مربع قائمة متغيرات الاختبار، ونقل متغير "الشركة" إلى مربع متغير التصنيف (Grouping Variable).



شكل 7.6: نافذة اختبار عينتين مستقلتين في الاختبارات اللامعلمية، لملف البيانات "المقابلات الشخصية".

بعد ذلك اضغط على خيار تعريف المجموعات (Define Groups) وأدخل القيم 1 و 2 في النافذة الفرعية التي ستظهر في مربعي المجموعة الأولى (Group 1) والمجموعة الثانية (Group 2) على التوالي، ثم اضغط استمرار.

ما قمنا به حتى الآن هو تعريف مشاهدات المتغير "نتائج المقابلات" كمتغير يضم مجموعتين أو عينتين مستقلتين، العينة الأولى هي نتائج مقابلات الشركة 1، والعينة الثانية هي نتائج مقابلات الشركة 2، والهدف سيكون اختبار ما إذا كانت كل من العينتين لهما نفس التوزيع أو اختبار ($H_0: \mu_1 = \mu_2$) باستخدام الاختبارات اللامعلمية لعينتين مستقلتين. في نافذة الاختبار، (شكل (7.6))، قم باختيار اختبار كولموجروف–سميرنوف Z إضافة لاختبار مان–ويتني U الافتراضي ثم اضغط موافق.

وفي نافذة المخرجات، سنحصل على نتائج الاختبارين، بالنسبة لاختبار مان–ويتني والذي سيظهر معه نتيجة اختبار ولكوكسون، (جدول (8.6))، ومن هذه النتائج، (P-value > 0.05)، نتجه لقبول الفرضية الصفرية القائلة بتساوي توزيع العينتين، بمعنى أن نتائج المقابلات في الشركتين متوافقة (متكافئة).

جدول 8.6: نتائج اختبار مان–وتني وولكوكسون لعينتين مستقلتين، للبيانات "المقابلات الشخصية".

Test Statistics[a]

	نتائج_المقابلات
Mann-Whitney U	195.500
Wilcoxon W	405.500
Z	-.122
Asymp. Sig. (2-tailed)	**.903**
Exact Sig. [2*(1-tailed Sig.)]	.904[b]

a. Grouping Variable: الشركة
b. Not corrected for ties.

ونلاحظ أن نفس النتيجة قد تم الوصول إليها من اختبار كولمجروف–سميرنوف، (جدول (9.6))، حيث أن القيمة الاحتمالية ($P\text{-value} > 0.05$).

جدول 9.6: نتائج اختبار كولموجروف–سميرنوف لعينتين مستقلتين، للبيانات "المقابلات الشخصية".

Test Statistics[a]

		نتائج_المقابلات
Most Extreme Differences	Absolute	.100
	Positive	.100
	Negative	-.100
Kolmogorov-Smirnov Z		.316
Asymp. Sig. (2-tailed)		**1.000**

a. Grouping Variable: الشركة

3.6 الاختبارات اللامعلمية لعدة عينات مستقلة

(Nonparametric Tests for Several Independent Samples)

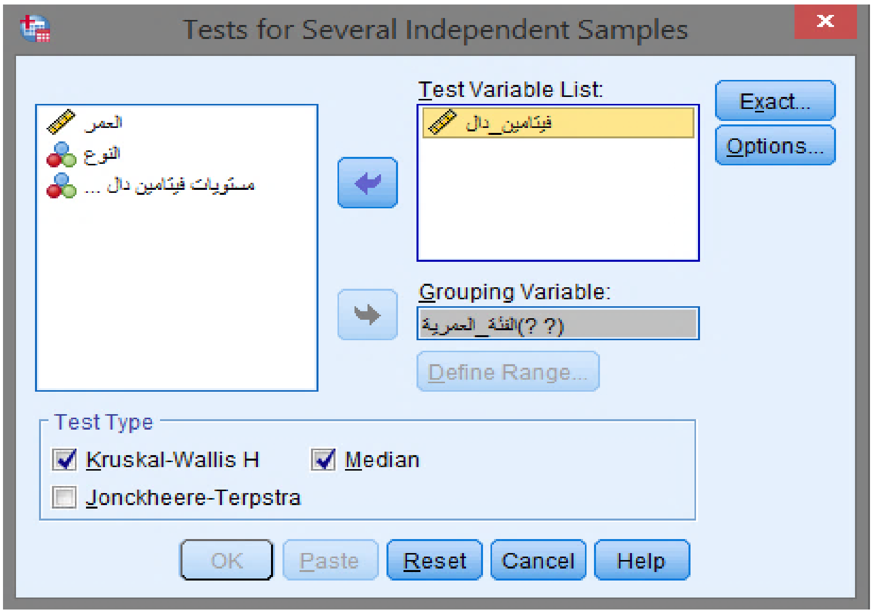
هذه الاختبارات تعادل في أهدافها اختبار تحليل التباين في اتجاه واحد، أي أنها تختبر تساوي عدة متوسطات، أو تكافؤ توزيع عدة متغيرات. ومن أهم الاختبارات اللامعلمية لعدة عينات مستقلة اختبار كروسكل–والس واختبار الوسيط.

1.3.6 اختبار كروسكل–والس H (Kruskal-Wallis H test) واختبار الوسيط (Median test)

يُعتبر اختبار كروسكل–والس امتدادا لاختبار مان–وتني U حيث أنه يختبر تساوي عدة متوسطات، أو يختبر ما إذا كانت العينات المستقلة هي مسحوبة من مجتمعات متطابقة التوزيع. وكذلك اختبار الوسيط الذي يتحقق من الاختلافات بين توزيعات المتغيرات محل الاختبار.

لتنفيذ هذه الاختبارات، قم بفتح ملف البيانات "فيتامين دال" في برنامج SPSS، حيث أننا سنقوم باختبار الفرضية الصفرية القائلة بأن معدلات مستويات فيتامين دال لدى الأشخاص هي متساوية في الفئات العمرية الثلاثة؛ (أقل من 30)، (من 30 إلى 50)، و(51 فأكثر)، أي أننا سنقوم باختبار الفرضية الصفرية $H_0: \mu_1 = \mu_2 = \mu_3$ بصورة لامعلمية.

في شريط الأوامر العلوي قم باختيار Analyze>Nonparametric Tests>Legacy Dialogs>k Independent Samples، فتظهر نافذة الاختبار المطلوبة، (شكل (8.6)).



شكل 8.6: نافذة اختبار عدة عينات مستقلة في الاختبارات اللامعلمية، (للبيانات "فيتامين دال").

في تلك النافذة، سنقوم باختيار المتغير الكمي "فيتامين_دال" ونقله لمربع قائمة متغيرات الاختبار، واختيار المتغير الوصفي "الفئة_العمرية" ونقله لمربع متغير التقسيم. بعد ذلك نضغط على أمر تعريف المدى (Define Range) والذي يتطلب إدخال القيمة الصغرى (Minimum) والقيمة الكبرى (Maximum) لتقسيمات المتغير الوصفي "الفئة_العمرية"، وهما القيمتان 1 و 3. ثم نضغط استمرار .

وبعد الضغط على موافق في نافذة الاختبار الرئيسية ستظهر النتائج في نافذة المخرجات. والنتيجة الأولى، (التي تظهر في جدول (10.6))، هي نتيجة اختبار كروسكال–والس ويظهر في الجدول الفرعي الأول عدد المشاهدات في كل تقسيم أو مجموعة، ومتوسط الرتب لكل مجموعة. وأما الجدول الفرعي الثاني فيشمل نتيجة الاختبار، وحيث أن القيمة الاحتمالية (P-value = 0.138) نستطيع الحكم بقبول الفرضية الصفرية $H_0: \mu_1 = \mu_2 = \mu_3$، بمعنى أننا نعتقد بأن مستويات فيتامين دال متكافئة في الفئات العمرية الثلاثة للأشخاص.

جدول 10.6: نتائج اختبار كروسكال–والس لثلاثة عينات مستقلة، للبيانات "فيتامين دال".

Ranks

فئات العمر		N	Mean Rank
فيتامين_دال	أقل من 30	16	32.16
	من 30 إلى 50	20	35.48
	51 فأكثر	24	25.25
	Total	60	

Test Statistics[a,b]

	فيتامين_دال
Chi-Square	3.954
df	2
Asymp. Sig.	**.138**

a. Kruskal Wallis Test
b. Grouping Variable: فئات العمر

وأما نتيجة اختبار الوسيط، (جدول (11.6))، فالنتيجة هي العكس، حيث أن القيمة الاحتمالية (P-value = 0.003)، مما يعني رفض الفرضية الصفرية، أي أنه بحسب اختبار الوسيط يوجد على الأقل متوسطان غير متساويان.

جدول 11.6: نتائج اختبار الوسيط لثلاثة عينات مستقلة، للبيانات "فيتامين دال".

Frequencies

		فئات العمر		
		أقل من 30	من 30 إلى 50	51 فأكثر
فيتامين_دال	> Median	9	14	5
	<= Median	7	6	19

Test Statistics[a]

	فيتامين_دال
N	60
Median	29.00
Chi-Square	11.401[b]
df	2
Asymp. Sig.	**.003**

a. Grouping Variable: فئات العمر
b. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 7.5.

وهذا الاختلاف في نتيجة الاختبارين يعود لاختلاف طبيعة حساب قيمة إحصاءة الاختبار لكل منهما، ويعود الحكم النهائي للباحث الذي قد يقوم بتنفيذ اختبارات مقارنة متعددة لكل متغيرين على حدة للحصول على فهم أعمق لتوزع المشاهدات في كل متغير.

4.6 الاختبارات اللامعلمية لعينتين مرتبطتين (Nonparametric Tests for Two Paired Samples)

كما هو الحال مع الاختبارات اللامعلمية المتعلقة بالعينات المستقلة، فإن اختبارات العينات المرتبطة لعينتين تختبر ما إذا كانت العينتين أو المتغيرين يتبعان نفس التوزيع. ويوفر برنامج SPSS أربعة اختبارات غير معلمية بالنسبة للعينات المرتبطة لاختبار الفرضية الصفرية $H_0: \mu_1 = \mu_2$. ويتم استخدام هذه الاختبارات بحسب طبيعة المتغيرات المراد اختبارها، وسنقوم باستعراضها فيما يلي:

1.4.6 اختبار الإشارة واختبار ولكوكسون للرتب ذات الإشارات

(Sign and Wilcoxon Signed-Rank Tests)

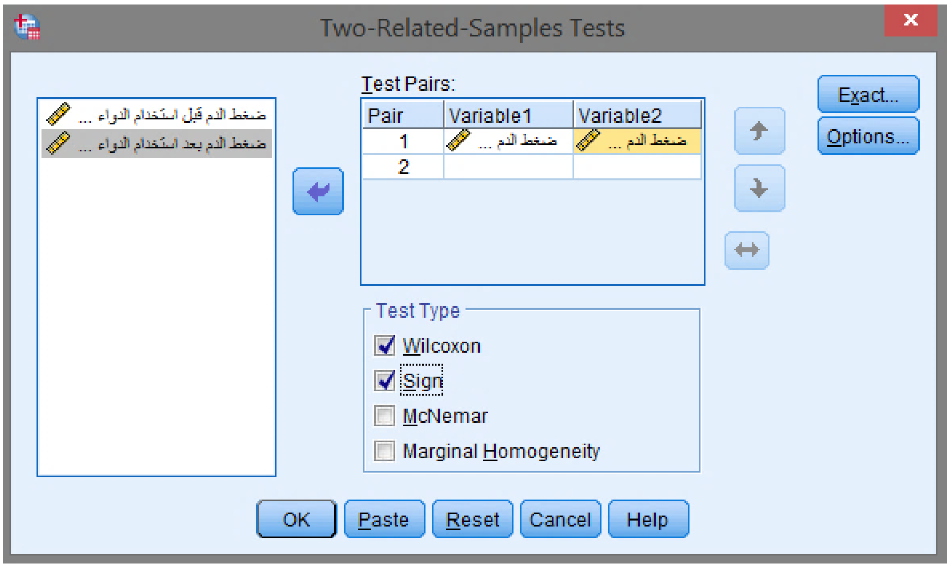
يتم استخدام هذين الاختبارين عند التعامل مع عينات مرتبطة كمية، حيث يقوم اختبار الإشارة بحساب الفروقات بين المتغيرين ثم تصنيفها إلى فروقات موجبة، سالبة، ومتعادلة (أي لا توجد فروقات). وإذا كان المتغيران لهما نفس التوزيع، فإن عدد الفروقات الموجبة والسالبة سيكون متقارب.

وأما اختبار ولكوكسون للرتب ذات الإشارات، فيُعد أفضل من اختبار الإشارة نظرا لأنه يأخذ في الاعتبار مقدار الفروقات بين المتغيرين إلى جانب إشارات الرتب، وبالتالي فإن إحصاءة الاختبار ستشمل على معلومات أكثر عن المتغيرين.

ولنقم الآن بتطبيق هذين الاختبارين في برنامج SPSS باستخدام ملف البيانات "عينات ضغط الدم"، حيث أننا سنختبر ما إذا كانت قياسات ضغط الدم للمرضى قبل تناول الدواء الموسع للأوردة مقاربة للقياسات بعد تناول الدواء. ونذكر القارئ هنا إلى أنه قد تم استخدام اختبار t في الفصل الرابع لاختبار نفس الفرضية الصفرية السابقة للمتغيرين بصورة معلمية؛ $H_0: \mu_D = 0$.

في شريط الأوامر العلوي قم باختيار Analyze>Nonparametric Tests>Legacy Dialogs>2 Related Samples، وفي نافذة الاختبار، الموضحة في شكل (9.6)، قم باختيار المتغير "عينة1" ونقله إلى مربع أزواج المتغيرات في الاختبار (Test Pairs) ثم اختيار المتغير "عينة2" ونقله إلى نفس المربع لتكوين زوج الاختبار.

بعد ذلك، وفي مربع نوع الاختبار (Test Type)، قم باختيار كل من اختبار ولكوكسون للرتب ذات الإشارات (Wilcoxon) واختبار الإشارة (Sign)، ثم اضغط موافق.



شكل 9.6: نافذة اختبار عينتين مرتبطتين في الاختبارات اللامعلمية، للمتغيرين "عينة1" و"عينة2"، (للبيانات "عينات ضغط الدم").

في نافذة المخرجات ستظهر نتيجة الاختبارين، ومن النتيجة الأولى، (الموضحة في جدول (12.6))، يمكن في الجدول الفرعي الأول ملاحظة متوسطات الرتب، (الموجبة، السالبة، والمتعادلة)، ومجموع هذه الرتب. وأما نتيجة الاختبار، (P-value = 0.474)، فتظهر في الجدول الفرعي الثاني، والتي تقترح قبول الفرضية الصفرية القائلة بعدم وجود فرق معنوي بين العينتين، بمعنى عدم وجود فرق في قياسات ضغط الدم قبل تناول الدواء وبعد تناوله.

جدول 12.6: نتائج اختبار ولكوكسون للرتب ذات الإشارات، للمتغيرين "عينة1" و"عينة2"، (للبيانات "عينات ضغط الدم").

Ranks

		N	Mean Rank	Sum of Ranks
ضغط الدم بعد استخدام الدواء - ضغط الدم قبل استخدام الدواء	Negative Ranks	5[a]	9.50	47.50
	Positive Ranks	10[b]	7.25	72.50
	Ties	0[c]		
	Total	15		

a. ضغط الدم بعد استخدام الدواء < ضغط الدم قبل استخدام الدواء
b. ضغط الدم بعد استخدام الدواء > ضغط الدم قبل استخدام الدواء
c. ضغط الدم بعد استخدام الدواء = ضغط الدم قبل استخدام الدواء

Test Statistics[a]

	ضغط الدم بعد استخدام الدواء - ضغط الدم قبل استخدام الدواء
Z	-.716-[b]
Asymp. Sig. (2-tailed)	**.474**

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

وهذه النتيجة هي مطابقة لما تم التوصل إليه لنفس المتغيرين باستخدام اختبار *t* في الفصل الرابع.

ومن النتيجة الثانية، (في جدول (13.6))، والخاصة بنتيجة اختبار الإشارة، نجد أن القيمة الاحتمالية تساوي (P-value = 0.302)، وبالتالي نقبل الفرضية الصفرية، وبالتالي نرى بأن نتيجة اختبار الإشارة متوافقة مع نتيجة اختبار ولكوكسون للرتب ذات الإشارات.

جدول 13.6: نتائج اختبار الإشارة للمتغيرين "عينة1" و"عينة2"، (للبيانات "عينات ضغط الدم").

Frequencies

		N
ضغط الدم بعد استخدام الدواء - ضغط الدم قبل استخدام الدواء	Negative Differences[a]	5
	Positive Differences[b]	10
	Ties[c]	0
	Total	15

a. ضغط الدم بعد استخدام الدواء < ضغط الدم قبل استخدام الدواء
b. ضغط الدم بعد استخدام الدواء > ضغط الدم قبل استخدام الدواء
c. ضغط الدم بعد استخدام الدواء = ضغط الدم قبل استخدام الدواء

Test Statistics[a]

	ضغط الدم بعد استخدام الدواء - ضغط الدم قبل استخدام الدواء
Exact Sig. (2-tailed)	**.302[b]**

a. Sign Test
b. Binomial distribution used.

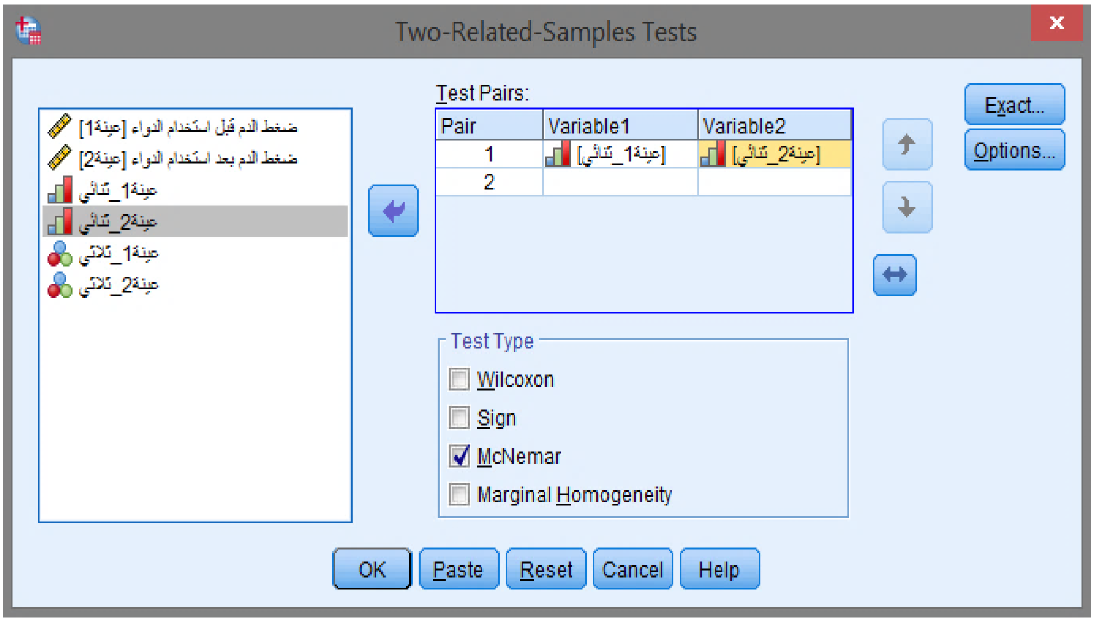
2.4.6 اختبار ماك نيمر (McNemar Test)

يُستخدم اختبار ماك نيمر مع المتغيرات الوصفية ثنائية القيم، حيث يقوم بتحديد ما إذا كان معدل الاستجابة الأولية (في المتغير الأول، أي قبل حدوث التغير) هو مساوي لمعدل الاستجابة النهائية (في المتغير الثاني، أي بعد حدوث التغير)، فإذا كانت نتيجة الاختبار تتضمن المساواة، فهذا يدل على عدم حدوث تغير في

الاستجابة من المتغير الأول إلى المتغير الثاني، أو بمعنى أبسط هذا يدل على عدم وجود فرق معنوي بين وسطي المجتمعين محل الدراسة.

لتطبيق هذا الاختبار، سنقوم أولا بتعريف متغيرين وصفيين ثنائيين جديدين في ملف البيانات السابق "عينات ضغط الدم"، هما تحويل من المتغيرين الكميين الأصليين "عينة1" و"عينة2" بحيث يأخذ كل من المتغير الأول الجديد، (وليكن اسمه "عينة1_ثنائي")، والمتغير الثاني الجديد، (وليكن اسمه "عينة2_ثنائي")، القيمة 1 للقيم 120 فأقل، والقيمة 2 للقيم الأكثر من 120. ويتم تعريف المتغيرين الجديدين باستخدام أمر إعادة ترميز في متغير جديد (Recode in Different Variables) في أمر تحوير (Transform) في شريط الأوامر العلوي كما وضحنا في الفصول السابقة.

الآن وفي شريط الأوامر العلوي بعد اختيار Analyze>Nonparametric Tests>Legacy Dialogs>2 Related Samples، نقوم باختيار المتغيرين "عينة1_ثنائي" و"عينة2_ثنائي" ونقلهما إلى مربع أزواج المتغيرات في الاختبار (Test Pairs)، ثم اختيار اختبار ماك نيمر فقط في مربع نوع الاختبار (Test Type)، كما يوضح الشكل (10.6)، ثم نضغط موافق.



شكل 10.6: نافذة اختبار ماك نيمر في الاختبارات اللامعلمية، للمتغيرين "عينة1_ثنائي" و"عينة2_ثنائي".

في نافذة المخرجات، (جدول (14.6))، سيحتوي الجدول الفرعي الأول على مصفوفة تصنيف القيم الثنائية في المتغيرين، ونلاحظ وجود حالة واحدة فقط مختلفة في التصنيف بين المتغيرين. أما نتيجة الاختبار في الجدول الفرعي الثاني فهي تبين عدم وجود تغير يُذكر قبل استخدام الدواء وبعده، (P-values = 1).

جدول 14.6: نتائج اختبار ماك نيمر للمتغيرين "عينة1_ثنائي" و"عينة2_ثنائي".

عينة1_ثنائي & عينة2_ثنائي

عينة1_ثنائي	عينة2_ثنائي	
	120 فأقل	أكثر من 120
120 فأقل	3	0
أكثر من 120	**1**	11

Test Statistics[a]

	عينة1_ثنائي & عينة2_ثنائي
N	15
Exact Sig. (2-tailed)	**1.000**[b]

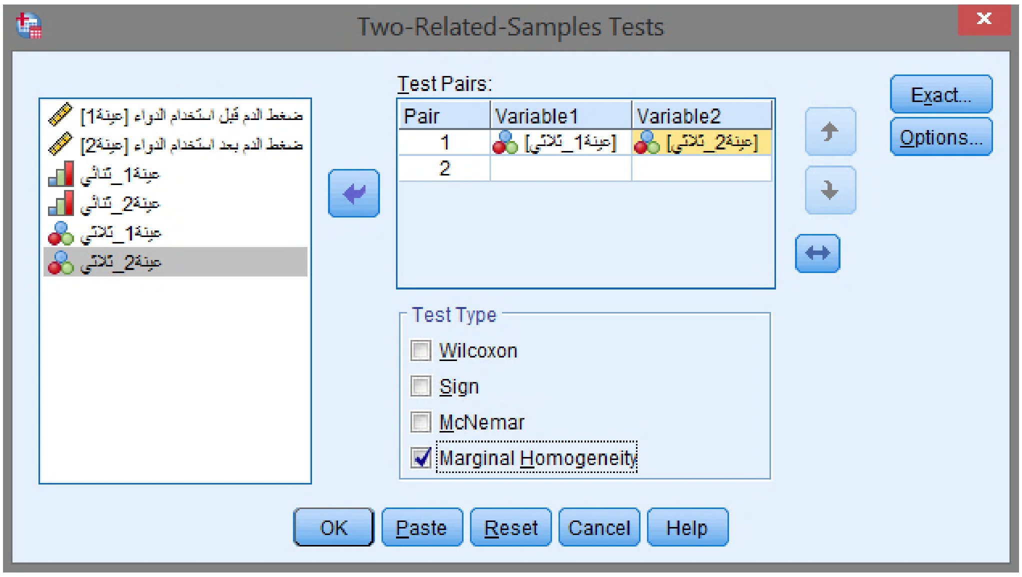
a. McNemar Test
b. Binomial distribution used.

3.4.6 اختبار التجانس الهامشي (Marginal Homogeneity Test)

يمكن اعتبار اختبار التجانس الهامشي للمتغيرات ذات القيم المتعددة، والذي يعتمد على توزيع مربع كاي، امتدادا لاختبار ماك نيمر للمتغيرات الثنائية، حيث أنه أيضا يبحث عن حدوث تغير في الاستجابة من المتغير الأول (قبل) إلى المتغير الثاني (بعد).

وكما قمنا في البند السابق بتعريف متغيرين ثنائيين جديدين، سنقوم هنا بتعريف متغيرين ثلاثيي (متعددي) القيمة، في ملف البيانات السابق "عينات ضغط الدم"، وسيكونا تحويل من المتغيرين الكميين الأصليين "عينة1" و"عينة2" بحيث يأخذ المتغير الأول المتعدد، (وليكن اسمه "عينة1_ثلاثي")، والمتغير الثاني المتعدد، (وليكن اسمه "عينة2_ثلاثي")، القيمة 1 للقيم 120 فأقل، والقيمة 2 للقيم من 121 إلى 140، والقيمة 3 للقيم الأكثر من 140. ويتم تعريف المتغيرين الجديدين باستخدام نفس أمر إعادة ترميز في متغير جديد في أمر تحوير (Transform) في شريط الأوامر العلوي.

الآن في شريط الأوامر العلوي بعد اختيار Analyze>Nonparametric Tests>Legacy Dialogs>2 Related Samples، قم باختيار المتغيرين "عينة1_ثلاثي" و"عينة2_ثلاثي" ونقلهما إلى مربع أزواج المتغيرات في الاختبار (Test Pairs)، ثم اختيار اختبار التجانس الهامشي فقط في مربع نوع الاختبار (Test Type)، كما يوضح الشكل (11.6)، ثم نضغط موافق.



شكل 11.6: نافذة اختبار التجانس الهامشي في الاختبارات اللامعلمية، للمتغيرين "عينة1_ثلاثي" و"عينة2_ثلاثي".

من الجدول (15.6)، الذي يحتوي على نتائج الاختبار في نافذة المخرجات، يتبين لنا من القيمة الاحتمالية للاختبار، (P-values = 0.317)، اتجاهنا لقبول الفرضية الصفرية القائلة بعدم حدوث تغير معنوي في قياسات ضغط الدم للمرضى قبل استخدام الدواء وبعد استخدامه.

جدول 15.6: نتائج اختبار التجانس الهامشي للمتغيرين "عينة1_ثلاثي" و"عينة2_ثلاثي".

Marginal Homogeneity Test

	عينة1_ثلاثي & عينة2_ثلاثي
Distinct Values	3
Off-Diagonal Cases	1
Observed MH Statistic	2.000
Mean MH Statistic	1.500
Std. Deviation of MH Statistic	.500
Std. MH Statistic	1.000
Asymp. Sig. (2-tailed)	**.317**

وخلاصة القول هنا، أن الاختبارات الأربعة اللامعلمية؛ اختبار الإشارة، اختبار ولكوكسون للرتب ذات الإشارات، اختبار ماك نيمر، واختبار التجانس الهامشي كلها أظهرت نفس النتيجة، فيما يتعلق بمثالنا المُستخدَم، باختلاف طبيعة المتغيرات المستخدمة في الاختبار؛ كمية، ثنائية، ومتعددة.

5.6 الاختبارات اللامعلمية لعدة عينات مرتبطة (Nonparametric Tests for Several Paired Samples)

كما هو الحال مع الاختبارات السابقة، فإن هذه الاختبارات اللامعلمية تقوم بمقارنة توزيع عينتين مرتبطتين أو أكثر، ويتوفر في برنامج SPSS ثلاثة اختبارات غير معلمية بالنسبة لـ k عينة مرتبطة تقوم باختبار الفرضية الصفرية $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$. وسنقوم باستعراض هذه الاختبارات فيما يلي:

1.5.6 اختبار فريدمان (Friedman test)

يمكن اعتبار اختبار فريدمان، أو اختبار فريدمان ANOVA كما يُطلق عليه أحيانا، اختبارا مكافئا لتحليل التباين في اتجاهين، حيث يقوم بإعطاء رتب من 1 إلى k للمتغيرات، (والتي عددها k)، وذلك بالنسبة لكل مشاهدة في العينة. فإذا كانت المتغيرات موجودة في الأعمدة، فهذا يعني أن الرتب ستوزع على المشاهدات في كل صف، ثم يتم حساب قيمة إحصاءة الاختبار، والتي تعتمد على توزيع مربع كاي، من هذه الرتب.

ولتطبيق اختبار فريدمان لنأخذ المثال التالي في جدول (16.6)، والذي يمثل أوزان 10 أشخاص (بالكيلوجرام) قاموا بتطبيق النظام الغذائي المعروف باسم الكيتو (Keto Diet)، هو نظام غذائي يعتمد على خفض نسبة الكربوهيدرات مقابل الاعتماد على عنصر الدهون، وذلك لمدة أربعة أشهر متواصلة.

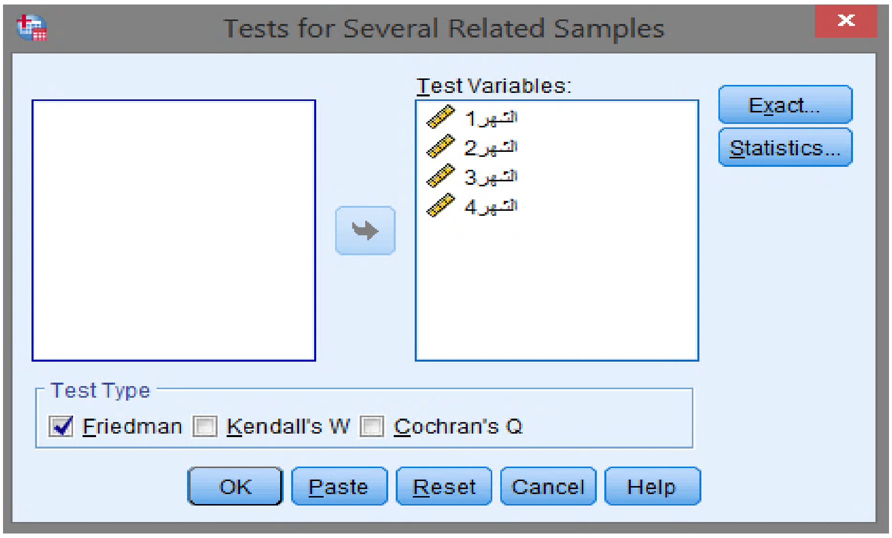
وبالتالي يمكن اعتبار أن أوزان هؤلاء الأشخاص في الأشهر الأربعة هي عينات أو متغيرات مرتبطة وليست مستقلة، لأن وزن الشخص في كل شهر يتعلق بالشهر السابق والشهر اللاحق خلال فترة تطبيق الحمية. وسيكون الهدف هنا هو استخدام اختبار فريدمان للعينات المرتبطة للتحقق من وجود تغير في أوزان هؤلاء الأشخاص خلال الأشهر الأربعة.

جدول 16.6: أوزان 10 أشخاص في أربعة أشهر أثناء تطبيق نظام الكيتو الغذائي.

الشهر1	الشهر2	الشهر3	الشهر4
75	90	84	86
75	83	80	81
74	89	95	98
69	80	75	77
65	73	70	75
68	77	80	75
70	79	86	80
71	79	80	82
64	81	83	81
73	91	88	90

سنبدأ بإدخال المتغيرات في جدول (16.6) في ملف بيانات جديد في برنامج SPSS باسم "نظام الكيتو"، بحيث يمثل كل شهر متغير، (بنفس الاسم).

الآن وفي شريط الأوامر العلوي قم باختيار Analyze>Nonparametric Tests>Legacy Dialogs>k Related Samples، وبعد ظهور نافذة الاختبار قم باختيار المتغيرات الأربعة ونقلها إلى مربع متغيرات الاختبار (Test Variables)، وستجد أن الخيار الافتراضي في النافذة هو اختبار فريدمان، كما يظهر في الشكل (12.6). قم بعدها بالضغط على موافق لتنفيذ الاختبار.



شكل 12.6: نافذة اختبار عدة عينات مرتبطة في الاختبارات اللامعلمية، لمتغيرات ملف البيانات "نظام الكيتو".

الجدول (17.6) يمثل النتائج التي ستظهر في نافذة المخرجات، ونلاحظ وجود قيم متوسطات الرتب لكل متغير في الجدول الفرعي الأول، أما الجدول الفرعي الثاني فتظهر فيه نتائج الاختبار ومنه نتجه لرفض الفرضية الصفرية $H_0: \mu_{\text{الشهر 1}} = \mu_{\text{الشهر 2}} = \mu_{\text{الشهر 3}} = \mu_{\text{الشهر 4}}$، وبالتالي يمكننا القول بأن اتباع نظام الكيتو الغذائي سيغير من الوزن بصورة معنوية (ملحوظة) خلال الأشهر المتتالية، (لكن هذا لا يعني بالضرورة التغير بنقصان الوزن).

جدول 17.6: نتائج اختبار فريدمان لمتغيرات ملف البيانات "نظام الكيتو".

Ranks

	Mean Rank
الشهر1	1.00
الشهر2	3.05
الشهر3	2.80
الشهر4	3.15

Test Statistics[a]

N	10
Chi-Square	18.576
df	3
Asymp. Sig.	**.000**

a. Friedman Test

2.5.6 اختبار كيندل W (Kendall's W test)

يمكن اعتبار إحصاءة اختبار كيندل W بمثابة **معامل للتوافق** (Coefficient of Concordance)، حيث أنه يستخدم تحديدا للتحقق من مدى اتفاق آراء المصوتين على موضوعات معينة. فمثلا، إذا تم سؤال مجموعة من الأشخاص (المشاهدات) عن رأيهم في أداء أربعة حكومات متعاقبة (والتي ستمثل المتغيرات) لأحدى الدول، عندها يكون اختبار كيندل W هو الاختبار الأنسب للتحقق من مدى توافق آراءهم.

وكما هو الحال مع الاختبارات اللامعلمية، فإن هذا الاختبار يعتمد على الرتب التي يتم تعيينها للقيم، حيث يقوم الاختبار بأخذ مجموع الرتب لكل متغير، وستتراوح قيمة إحصاءة كيندل W ما بين القيمة 0، (والتي تعني عدم وجود أي توافق في الآراء)، والقيمة 1، (والتي تعني التوافق التام في الآراء).

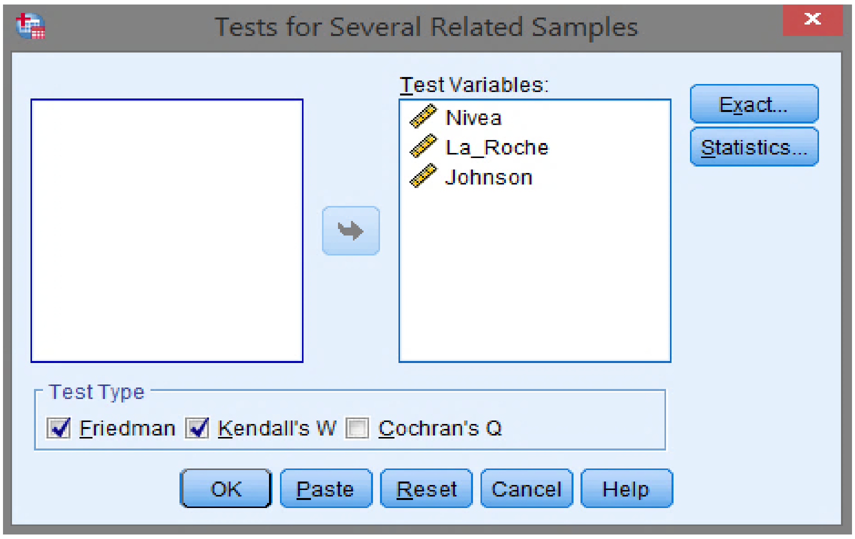
ولنأخذ المثال التالي في جدول (18.6)، والذي يمثل تقييمات 12 سيدة تم سؤالهن عن آراءهن حول مستحضرات لتفتيح البشرة من ثلاثة شركات شهيرة هي Nivea، La Roche، وJohnson، فقمن بتقييم كل مستحضر على مقياس من 1 إلى 10، (أي من التقييم الأسوأ إلى التقييم الأفضل على الترتيب).

جدول 18.6: آراء 12 سيدة حول ثلاثة مستحضرات لتفتيح البشرة، (من 1 إلى 10).

Nivea	La Roche	Johnson
3	9	4
5	8	3
2	9	2
4	10	3
1	7	1
6	8	1
7	9	3
1	10	5
2	10	6
5	8	8
5	7	10
6	8	9

وسيكون الهدف هنا هو التحقق من مدى توافق آراء هذه العينة من السيدات حول الشركات الثلاثة. وسنقوم بإدخال المتغيرات في جدول (18.6) في ملف بيانات جديد في برنامج SPSS باسم "شركات التجميل"، بحيث تمثل كل شركة متغير وبنفس الاسم.

في شريط الأوامر العلوي قم باختيار Analyze>Nonparametric Tests>Legacy Dialogs>k Related Samples، وبعد ظهور نافذة الاختبار قم باختيار المتغيرات الثلاثة ونقلها إلى مربع متغيرات الاختبار (Test Variables)، كما يظهر في الشـــكل (13.6). قم باختيار اختبار كيندل W، (ويمكننا إبقاء خيار اختيار اختبار فريدمان موجودا لمقارنة النتائج)، ونقوم بعدها بالضغط على موافق.



شكل 13.6: نافذة اختبار عدة عينات مرتبطة في الاختبارات اللامعلمية، لمتغيرات ملف البيانات "شركات التجميل".

وفي نافذة المخرجات، سنبدأ بالتعليق على نتيجة اختبار كيندال W أولا، والتي تُعرض في جدول (19.6)، ونلاحظ من نتيجة الاختبار في الجدول الفرعي الثاني أن معامل أو إحصاءة كندل للتوافق تساوي 0.539 مما يدل على

وجود اختلاف في آراء السيدات في العينة حول مستحضرات تفتيح البشرة في الشركات الثلاثة، ويؤكد هذا الاستنتاج القيمة الاحتمالية للاختبار (P-value = 0.002). وأما الجدول الأول الفرعي في جدول (19.6) فيمثل متوسطات الرتب في المتغيرات (الشركات) الثلاثة.

جدول 19.6: نتائج اختبار كيندل W لمتغيرات ملف البيانات "شركات التجميل".

Ranks

	Mean Rank
Nivea	1.42
La_Roche	2.79
Johnson	1.79

Test Statistics

N	12
Kendall's W[a]	**.539**
Chi-Square	12.933
df	2
Asymp. Sig.	**.002**

a. Kendall's Coefficient of Concordance

ومن الجدول الفرعي الثاني، في جدول (20.6)، نرى بأننا نتجه لرفض الفرضية الصفرية القائلة بتساوي أو توافق آراء السيدات؛ $H_0: \mu_{Nivea} = \mu_{La_Roche} = \mu_{Johnson}$، وبالتالي يمكن رؤية توافق الاستنتاجات لكلا الاختبارين.

جدول 20.6: نتائج اختبار فريدمان لمتغيرات ملف البيانات "شركات التجميل".

Ranks

	Mean Rank
Nivea	1.42
La_Roche	2.79
Johnson	1.79

Test Statistics[a]

N	12
Chi-Square	12.933
df	2
Asymp. Sig.	**.002**

a. Friedman Test

3.5.6 اختبار كوكران Q (Cochran's Q test)

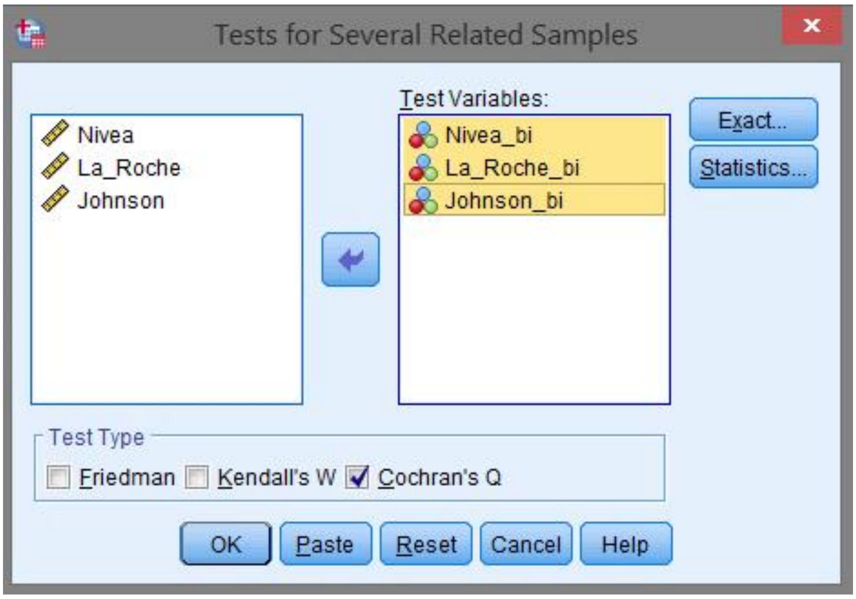
يُعد اختبار كوكران Q مطابقا لاختبار فريدمان، إلا أنه يتعامل مع قيم المتغيرات الثنائية، ويمكن أيضا النظر إلى هذا الاختبار على أنه امتداد لاختبار ماك نيمر عند التعامل مع k عينة. ولتطبيق هذا الاختبار، لنفرض أن آراء السيدات في المثال السابق، (في جدول (18.6))، قد تم تسجيلها بإجابات جيد أو سيئ على مستحضرات الشركات الثلاثة، أي أنها أصبحت إجابات ثنائية، كما يظهر في الجدول (21.6).

عندها سنقوم بإدراج المتغيرات الثلاثة في نفس ملف البيانات السابق "شركات التجميل" بالأسماء الجديدة التي تظهر في الجدول، بحيث يتم تعيين القيمة 1 للإجابة "سيئ"، والقيمة 2 للإجابة "جيد"، مع مراعاة تعريف المشاهدات في عامود القيم (Values) في نافذة عرض المتغيرات (Variable View) في ملف البيانات.

جدول 21.6: آراء 12 سيدة حول ثلاثة مستحضرات لتفتيح البشرة، (جيد/سيئ).

Johnson_bi	La_Roche_bi	Nivea_bi
سيئ	جيد	سيئ
سيئ	جيد	سيئ
سيئ	جيد	سيئ
سيئ	جيد	سيئ
سيئ	جيد	سيئ
سيئ	جيد	جيد
سيئ	جيد	جيد
سيئ	جيد	سيئ
جيد	جيد	سيئ
جيد	جيد	سيئ
جيد	جيد	سيئ
جيد	جيد	جيد

في شريط الأوامر العلوي سنقوم باختيار Analyze >Nonparametric Tests> Legacy Dialogs> k Related Samples، وفي نافذة الاختبار نقوم باختيار المتغيرات الثلاثة؛ Nivea_bi، La_Roche_bi، وJohnson_bi ونقلها إلى مربع متغيرات الاختبار (Test Variables)، كما يظهر في الشكل (14.6). ثم نقوم باختيار اختبار كوكران Q (Cochran's Q)، ثم نضغط على موافق.



شكل 14.6: نافذة اختبار عدة عينات مرتبطة في الاختبارات اللامعلمية، للمتغيرات الثنائية.

وفي نافذة المخرجات، سنحصل على النتائج المطلوبة، والتي يمثلها جدول (22.6)، ويوضح الجدول الفرعي الأول تكرار الإجابات بسيئ (1) وجيد (2) للشركات الثلاثة، ولاحظ منها أن الشركة (La_Roche) لم تحصل على أي رأي "سيئ" في عينة السيدات.

جدول 22.6: نتائج اختبار فريدمان لمتغيرات ملف البيانات "شركات التجميل".

Frequencies

	Value	
	1	2
Nivea_bi	9	3
La_Roche_bi	0	12
Johnson_bi	8	4

Test Statistics

N	12
Cochran's Q	13.273[a]
df	2
Asymp. Sig.	**.001**

a. 1 is treated as a success.

وتُظهر نتيجة اختبار كوكران Q في الجدول الفرعي الثاني وجود اختلاف في آراء السيدات حول مستحضرات التجميل في الشركات الثلاثة، حيث أن القيمة الاحتمالية للاختبار (P-value = 0.001)، تقودنا لرفض الفرضية الصفرية $H_0: \mu_{Nivea_bi} = \mu_{La_Roche_bi} = \mu_{Johnson_bi}$.

6.6 تحليل الموثوقية (Reliability Analysis)

في كثير من الدراسات قد يصعب على الباحث التأكد من أن بيانات الدراسة المستخدمة في التحليل الإحصائي هي بيانات تمثل الواقع الحقيقي، وخاصة تلك البيانات التي تعتمد على إجابات الأشخاص عن مجموعة من الأسئلة المتعلقة بآرائهم وأشهرها أسئلة **الاستبيانات** (Questionnaires)، والتي يرتكز معظمها على أسئلة ليكارت (Likert) المتعددة؛ الثنائية، الثلاثية، الخماسية، والسباعية.

إذ أن المجيب أو المستهدف من الاستبيان قد يُعطي إجابات (بيانات) غير صحيحة وذلك لعدة أسباب منها عدم معرفته بإجابة السؤال أو الفقرة فيجيب بالتخمين، أو قد لا يركز في الأسئلة جيدا أو قد لا يفهم التعليمات فهما صحيحا، أو قد يخشى قول الصدق، أو قد يشعر أن السؤال أو الفقرة شخصي جدًا في طبيعته، وغيرها من الأسباب، لذلك لابد في كثير من الدراسات التأكد من مدى مصداقية الإجابات على الاستبيان وتقدير أخطاء الاستجابة.

وتوجد عدة طرق لقياس مدى صدق وثبات الاستبيان، وسنتطرق لأهم ثلاثة منها؛

1.6.6 **معامل ثبات كرونباخ الفا** (Cronbach's Alpha Coefficient)

يعد معامل ثبات كرونباخ ألفا من أشهر مقاييس **الثبات الداخلي** (Internal Consistency Reliability) للاستبيان، ويعتمد على حساب الاختلافات (التباينات) الداخلية بين إجابات الأسئلة في الاستبيان من خلال استخدام الصيغة التالية:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} S_i^2}{S^2}\right)$$

حيث

k = عدد أسئلة الاستبيان، S_i^2 = التباين لإجابات السؤال i ، S^2 = التباين لإجابات جميع الأسئلة. وارتفاع قيمة المعامل يُعتبر مؤشرا على ارتفاع درجة ثبات أو موثوقية إجابات المستهدفين على الاستبيان.

2.6.6 **معامل ثبات التجزئة النصفية** (Split-Half Reliability Coefficient)

يعتمد حساب هذا المعامل على تجزئة أسئلة الاستبيان (المتغيرات) إلى جزأين متساويين، ثم حساب الارتباط بين مجاميع المتغيرات في كل جزء، (وليكن r)، ثم حساب معامل الثبات أو الموثوقية ويُعرف أيضا **بمعامل سبيرمان–براون** (Spearman-Brown) كالتالي:

$$Spearman - Brown\ Coefficient = \frac{2\,r}{1+r}$$

وكلما كانت قيمة معامل الثبات أكبر كلما دل ذلك على ثبات وصدق إجابات المستهدفين.

3.6.6 **معامل ثبات جوتمان** (Guttman Coefficient)

يعتمد هذا المعامل أيضا على تجزئة أسئلة الاستبيان إلى نصفين متساويين ثم حساب تباين الإجابات في كل نصف على حدة، S_1^2 وS_1^2، ثم يتم حساب التباين الكلي لجميع إجابات الاستبيان S^2، وتكون صيغة حساب معامل جوتمان للثبات معرفة بالصورة التالية:

$$Guttman\ Coefficient = 2\left(1 - \frac{S_1^2 + S_2^2}{S^2}\right)$$

ويمكن اعتبار معامل جوتمان حالة خاصة من معامل كرونباخ ألفا، والقيمة المرتفعة للمعامل تدل على ثبات أفضل للإجابات على الاستبيان.

لنأخذ الآن المثال التالي لاستبيان افتراضي يتضمن 10 أسئلة تم توزيعه على عينة مكونة 100 شخص، حيث تمثل استجابة هؤلاء الأشخاص للأسئلة التالية المتعلقة بجودة المنتجات الغذائية العربية والأجنبية في العموم:

س1:	المنتجات الغذائية العربية كلها متشابهة.
س2:	يفضل دائما شراء المنتجات الغذائية الأجنبية.
س3:	المنتجات الغذائية العربية لا تتوفر فيها الجودة القياسية.
س4:	المنتجات الغذائية الأجنبية دائما ممتازة في جودتها.
س5:	المنتجات الغذائية العربية سعرها مبالغ فيه.
س6:	المنتجات الغذائية العربية تحتوي دائما على مواد حافظة ضارة.
س7:	مصانع المنتجات الغذائية العربية لا تهتم بآراء المستهلكين.
س8:	المنتجات الغذائية الأجنبية ليست دائما الأجود.
س9:	المنتجات الغذائية الأجنبية تفتقد لذوق المستهلك العربي.
س10:	يجب تقليص استيراد المنتجات الغذائية الأجنبية.

وإجابة كل سؤال كانت على مقياس من 1 إلى 9، بحيث أن القيمة 1 تعني غير موافق على الإطلاق تدرجا في رأي الشخص إلى القيمة 9 والتي تعني موافق بشدة. وإجابات المستهدفين من هذا الاستبيان هي مبينة في الجدول التالي، جدول (23.6)، والتي سنقوم بإدراجها في ملف بيانات جديد في برنامج SPSS باسم "استبيان المنتجات الغذائية". وسيكون الهدف هو استخدام بعض مؤشرات تحليل الموثوقية للتحقق من مدى موثوقية أو ثبات إجابات العينة المستهدفة من الاستبيان. وسنقوم بكتابة أسئلة الاستبيان في العامود الخاص بتعريف المتغيرات (Label) في ملف البيانات لمزيد من التوضيح أثناء تفسير النتائج التي سنحصل عليها لاحقا من تحليل الموثوقية.

جدول 23.6: إجابات عينة من 100 شخص على 10 أسئلة في استبيان حول جودة المنتجات الغذائية العربية والأجنبية.

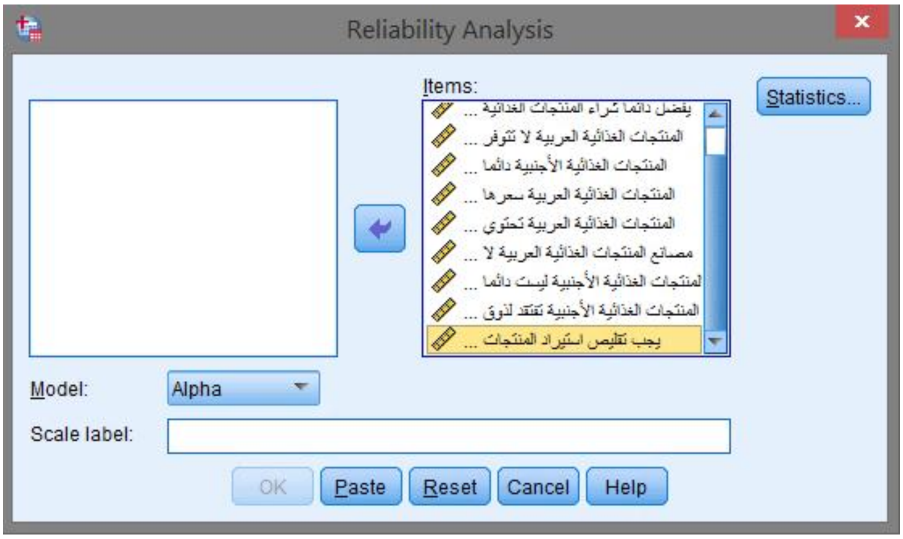
س1	س2	س3	س4	س5	س6	س7	س8	س9	س10
7	7	5	7	4	3	8	9	9	7
3	4	6	3	4	4	4	4	4	4
3	5	5	4	5	4	5	6	6	7
6	6	4	4	5	5	7	6	7	5
3	3	4	4	7	3	3	4	2	5
4	5	5	5	4	4	4	5	4	5
4	3	4	2	3	4	4	3	3	4
6	6	4	5	4	5	5	6	7	7
5	4	5	5	7	2	5	5	5	5
5	5	6	5	6	3	4	5	6	4
2	2	4	3	6	6	3	3	3	4
5	4	5	5	2	6	5	3	5	5
4	7	7	5	4	7	6	4	6	4
4	7	4	6	7	3	4	7	5	3
6	6	6	4	3	6	6	7	6	4
5	5	5	4	4	4	4	5	5	5
5	4	6	4	6	5	4	5	4	5
5	3	4	6	3	5	5	7	5	6

س1	س2	س3	س4	س5	س6	س7	س8	س9	س10
5	6	5	7	5	8	5	7	7	4
4	3	4	5	5	4	4	4	2	4
4	4	3	5	7	5	3	5	5	3
4	3	6	5	4	2	5	3	6	4
4	6	7	5	5	3	4	5	5	2
5	6	6	8	4	5	6	7	7	7
5	5	5	5	6	6	5	5	5	5
4	3	2	5	5	5	3	2	4	3
4	4	5	3	7	4	4	5	5	4
4	3	5	3	1	3	4	5	6	4
6	6	7	6	5	5	7	7	4	6
2	4	3	1	4	4	5	4	2	1
4	3	3	3	6	6	5	4	4	4
5	6	5	4	3	3	5	5	3	6
7	5	7	7	5	4	6	2	4	5
4	3	2	3	4	4	4	3	5	3
4	5	5	4	3	4	5	5	4	3
4	6	5	2	4	5	6	5	5	6

(تابع) جدول 23.6: إجابات عينة من 100 شخص على 10 أسئلة في استبيان حول جودة المنتجات الغذائية العربية والأجنبية.

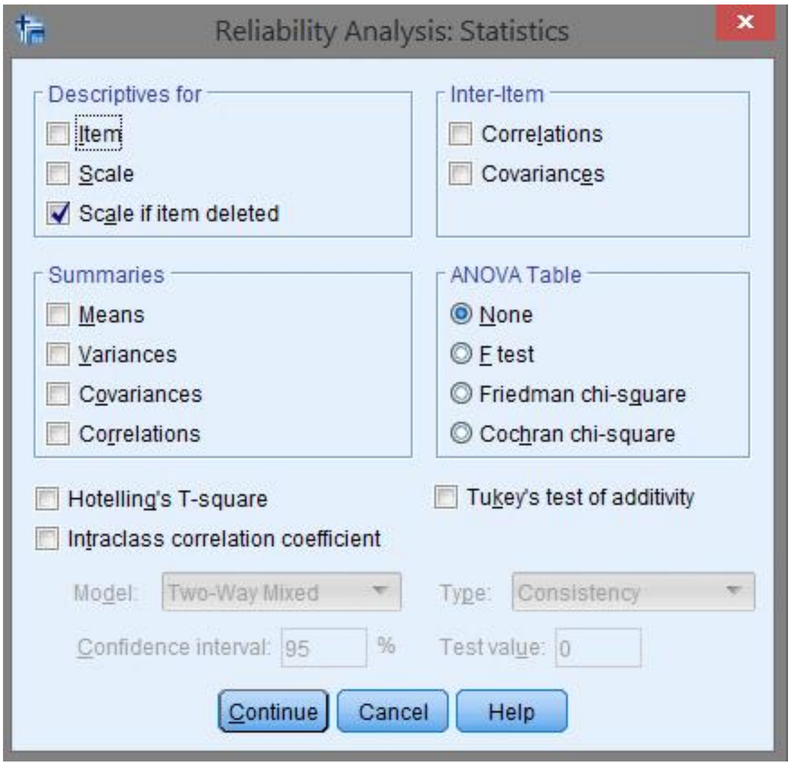
6	6	5	6	5	6	6	5	7	8	5	3	3	3	4	6	3	2	3	5
5	4	5	5	4	8	6	5	5	4	4	5	5	5	3	4	4	5	5	5
5	3	4	3	5	6	3	4	5	3	3	5	5	5	3	7	5	6	6	5
4	5	5	5	3	3	4	4	3	4	5	5	5	4	4	4	4	4	9	5
5	4	5	2	6	5	4	6	4	4	3	4	4	6	3	3	3	3	4	2
6	6	4	5	4	4	4	6	3	5	2	3	5	5	5	3	1	6	5	3
4	4	4	4	4	5	6	5	3	4	2	2	5	4	5	3	4	4	3	2
5	7	4	5	5	7	7	6	3	4	5	4	6	4	7	7	5	4	5	4
6	4	5	6	3	3	4	3	6	5	6	6	7	5	5	4	4	4	6	5
6	6	6	4	8	5	6	6	4	5	3	4	6	7	5	4	5	6	6	5
3	5	4	3	3	5	6	4	4	5	3	5	3	3	5	5	5	5	4	5
7	4	5	4	7	5	7	5	4	4	2	4	3	3	7	3	4	4	3	3
1	2	2	3	3	2	2	2	3	1	5	5	4	5	1	5	3	6	4	6
7	6	7	7	4	4	5	5	6	5	6	7	6	5	5	8	3	5	6	6
5	5	4	6	6	5	5	6	6	6	4	4	5	6	8	3	3	5	3	5
6	7	7	4	6	3	8	7	5	6	6	5	6	4	4	4	5	6	5	6
5	6	6	4	4	7	6	6	5	6	2	3	3	2	6	2	1	4	3	3
3	5	5	3	5	3	4	4	6	3	5	6	4	5	3	4	6	6	5	5
7	6	7	6	6	6	6	7	6	4	7	5	6	6	5	3	7	5	4	5
3	4	5	2	3	6	4	4	3	4	5	4	4	5	5	5	5	6	4	8
4	4	3	3	4	5	2	1	4	1	4	5	5	5	4	3	6	4	6	4
6	6	5	5	3	6	5	4	7	5	3	4	4	4	5	7	6	4	3	3
4	5	5	5	6	3	6	4	4	5	2	4	4	5	4	4	3	6	5	5
4	5	4	3	2	4	5	5	5	5	4	4	5	4	3	6	4	3	2	2
4	5	2	3	3	2	3	4	2	4	6	6	3	7	6	3	4	5	6	7
6	7	7	5	5	6	6	5	5	5	4	6	6	4	4	3	6	5	6	4
6	6	3	3	5	7	4	6	5	6	6	5	5	5	6	6	4	7	6	3
6	6	6	5	5	4	5	5	6	5	3	5	5	5	5	6	5	5	3	4
3	4	2	2	4	7	3	3	3	3	6	5	4	6	1	4	5	5	4	4
4	4	2	4	5	3	5	2	4	4	5	3	5	5	6	6	4	3	4	2
8	7	6	5	3	2	7	5	4	5	3	4	4	5	6	6	4	5	4	5
1	4	1	5	4	4	3	3	5	3	6	6	7	5	7	5	6	7	6	6

في شريط الأدوات العلوي، قم باختيار Analyze>Scale>Reliability Analysis فتظهر نافذة التحليل كما هو موضح في الشكل (15.6). قم أولا باختيار كل المتغيرات العشرة ونقلها إلى مربع المواد أو المتغيرات (Items).



شكل 15.6: نافذة تحليل الموثوقية لمتغيرات ملف البيانات "استبيان المنتجات الغذائية".

بعد ذلك قم بالضغط على خيار الإحصاءات (Statistics) في أعلى اليمين فتظهر النافذة الفرعية الخاصة به، كما يوضح الشكل (16.6). وفي تلك النافذة قم باختيار خيار المقياس عند حذف المتغير (Scale if item deleted) في مربع الإحصاءات الوصفية (Descriptive for).



شكل 16.6: نافذة اختيار الإحصاءات في تحليل الموثوقية.

ونلفت نظر المستخدم هنا إلى إمكانية اختيار عرض عدة نتائج مثل مصفوفة الارتباطات والتغايرات من مربع العلاقات بين المتغيرات (Inter-Item)، وعرض المتوسطات والتباينات وغيرها من مربع الملخصات (Summaries).

سنكتفي بخيارنا الحالي الآن ونضغط استمرار للعودة للنافذة الرئيسية، وفي تلك النافذة، سنبقي على الخيار الافتراضي لطريقة حساب معامل الموثوقية في مربع النموذج (Model)، وهو معامل كرونباخ ألفا، ثم نضغط موافق فتظهر النتائج المطلوبة في نافذة المخرجات.

والجدول الأول يوضح عدد المشاهدات المستخدمة والمستثناة، والجدول الثاني في نافذة المخرجات، (جدول (24.6))، يعرض قيمة معامل كرونباخ ألفا في الجدول الفرعي الأول وهي قيمة جيدة ($\alpha = 0.799$) تدل على ثبات الإجابات في الاستبيان.

جدول 24.6: نتائج معامل كرونباخ ألفا وإحصاءات المتغيرات لملف بيانات "استبيان المنتجات الغذائية".

Reliability Statistics

Cronbach's Alpha	N of Items
.799	10

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
المنتجات الغذائية العربية كلها متشابهة	41.70	53.889	.655	.759
يفضل دائما شراء المنتجات الغذائية الأجنبية	41.46	55.625	.670	.760
المنتجات الغذائية العربية لا تتوفر فيها الجودة القياسية	41.50	56.879	.549	.773
المنتجات الغذائية الأجنبية دائما ممتازة في جودتها	41.72	58.204	.485	.780
المنتجات الغذائية العربية سعرها مبالغ فيه	41.61	66.261	**.060**	**.829**
المنتجات الغذائية العربية تحتوي دائما على مواد حافظة ضارة	41.65	64.210	**.143**	**.820**
مصانع المنتجات الغذائية العربية لا تهتم بآراء المستهلكين	41.55	56.028	.587	.768
المنتجات الغذائية الأجنبية ليست دائما الأجود	41.42	54.973	.619	.764
المنتجات الغذائية الأجنبية تفتقد لذوق المستهلك العربي	41.53	56.716	.514	.776
يجب تقليص استيراد المنتجات الغذائية الأجنبية	41.75	55.927	.568	.770

وأما الجدول الفرعي الثاني، والذي يمثل بعض المقاييس الإحصائية لمتغيرات (أسئلة) الاستبيان فيمكن الاستفادة منه بصورة خاصة للتحقق من الأسئلة (أو بمعنى آخر إجابات المستهدفين) التي وجودها يقلل من مدى ثبات الاستبيان، وهذه الأسئلة هي التي ترتفع قيمة معامل كرونباخ ألفا بحذفها، وهذا يمكن مراقبة في العامود الأول إلى اليمين من هذا الجدول (Cronbach's Alpha if Item Deleted).

ونلاحظ أن السؤالين الخامس (المنتجات الغذائية العربية سعرها مبالغ فيه) والسادس (المنتجات الغذائية العربية تحتوي دائما على مواد حافظة ضارة) إذا ما تم حذفهما فإن قيمة معامل كرونباخ ألفا سوف ترتفع بصورة ملحوظة؛ (0.829) و(0.820) على الترتيب، إلى جانب كون معاملي الارتباط بين هذين السؤالين والأسئلة ككل هما الأدنى مقارنة بباقي معاملات الارتباط في العامود الثاني إلى اليمين، (0.060) و(0.143) على الترتيب.

لذلك قد يكون من الموصى به أحيانا حذف هذين السؤالين، (أو إحداهما على الأقل)، إذا ما رغب الباحث برفع مستوى موثوقية أو ثبات الاستبيان، وبالتالي رفع القيمة العلمية للدراسة ككل.

من جديد، يمكننا استخدام معامل ثبات التجزئة النصفية ومعامل جوتمان ضمن التجزئة النصفية ومعامل كرونباخ ألفا ضمن التجزئة النصفية لمزيد من التحقق حول مدى ثبات أسئلة الاستبيان.

في شريط الأدوات العلوي، قم من جديد باختيار Analyze>Scale>Reliability Analysis، وفي نافذة التحليل قم باختيار كل المتغيرات العشرة ونقلها إلى مربع المواد أو المتغيرات (Items)، وهنا يمكننا الاستغناء عن اختيار خيار المقياس عند حذف المتغير (Scale if item deleted) في خيار الإحصاءات (Statistics) حيث أنه قد تم التطرق إليه في النتيجة السابقة. نقوم بعدها باختيار طريقة الثبات في التجزئة النصفية (Split-half) كطريقة حساب معامل الموثوقية في مربع النموذج (Model)، ثم نضغط موافق.

وفي نافذة المخرجات، من الجدول الثاني، والموضح هنا في جدول (25.6)، نلاحظ قيم معاملات الموثوقية وهي معروضة بالترتيب التالي، (حيث يمثل الجزء الأول (Part 1) الخمسة أسئلة الأولى في الاستبيان، والجزء الثاني (Part 2) الخمسة أسئلة المتبقية من السؤال السادس وحتى العاشر)؛

- قيم معامل كرونباخ ألفا للجزء الأول والثاني من المتغيرات وهي (0.662) و(0.673) على الترتيب ونلاحظ أنها أضعف من قيمة المعامل ألفا للأسئلة ككل، (جدول (24.5)).
- أما معامل الارتباط بين مجاميع المتغيرات في كل جزء فهو (0.659) ويُعد مقبولا إلى حد ما، وأما معامل معامل ثبات التجزئة النصفية أو معامل سبيرمان–براون فهو (0.794) ويُعتبر جيدا.
- والمعامل الأخير في الجدول وهو معامل جوتمان، (0.794)، يُعد أيضا جيدا.

جدول 25.6: نتائج طريقة الثبات في التجزئة النصفية لمتغيرات ملف البيانات "استبيان المنتجات الغذائية".

Reliability Statistics

Cronbach's Alpha	Part 1	Value	**.662**
		N of Items	5[a]
	Part 2	Value	**.673**
		N of Items	5[b]
	Total N of Items		10
Correlation Between Forms			**.659**
Spearman-Brown Coefficient	Equal Length		**.794**
	Unequal Length		.794
Guttman Split-Half Coefficient			**.794**

a. The items are: المنتجات الغذائية العربية كلها متشابهة، يفضل دائما شراء المنتجات الغذائية الأجنبية، المنتجات الغذائية العربية لا تتوفر فيها الجودة القياسية، المنتجات الغذائية الأجنبية دائما ممتازة في جودتها, المنتجات الغذائية العربية سعرها مبالغ فيه.

b. The items are: مصانع المنتجات الغذائية العربية لا تهتم بآراء المستهلكين، المنتجات الغذائية الأجنبية ليست دائما الأجود، المنتجات الغذائية الأجنبية تفتقد لذوق المستهلك العربي، يجب تقليص استيراد المنتجات الغذائية الأجنبية, المنتجات الغذائية العربية تحتوي دائما على مواد حافظة ضارة.

وهكذا نخلص إلى أن الإجابات على أسئلة الاستبيان بالنسبة لهذه العينة المكونة من 100 شخص تُعد جيدة الثبات أو الموثوقية من خلال نتائج المقاييس السابقة، مع الأخذ بالاعتبار حذف السؤال الخامس أو السادس أو كلاهما إذا ما أردنا رفع درجة ثبات أو موثوقية الاستبيان.

الملحق

ملفات البيانات المستخدمة في الكتاب

م	اسم ملف البيانات في SPSS	تعريف ملف البيانات	الصفحة
.1	**مثال1**	بيانات خاصة بعشرة أطفال في إحدى المؤسسات التعليمية في مرحلتي رياض الأطفال والابتدائية، وتشمل ثلاثة متغيرات هي وزن الطفل (بالكجم) وعمره (بالسنوات) وجنسه.	9
.2	**بيانات الطلبة**	المشاهدات الخاصة بـ 35 طالبا جامعيا مقاسة لثمانية متغيرات هي: درجة الطالب في المقرر1 (من 100 درجة)، درجة الطالب في المقرر2، درجة الطالب في المقرر3، عمر الطالب (بالسنوات)، جنس الطالب، ترتيب الفصل الدراسي للطالب، عدد أفراد أسرة الطالب، وعدد الحجرات في منزل الطالب.	18
.3	**مثال2**	بيانات المثال1 مصنفة بحسب جنس الطالب.	26
.4	**مثال3**	بيانات المثال1 مصنفة بحسب جنس الطالب ثم عمره.	26
.5	**تطبيقات التواصل الاجتماعي**	آراء عينة مكونة من 12 شخص ردا على المطلوب التالي: "ضع إشارة (√) أمام التطبيقات التي تستخدمها وإشارة (×) أمام التطبيقات التي لا تستخدمها". وهذه التطبيقات هي؛ فيسبوك، واتساب، فايبر، وإنستجرام.	27
.6	**مثال4**	بيانات تمثل التقدير الدراسي والقسم العلمي لعينة من 15 طالب.	37
.7	**أسعار النفط**	أسعار النفط العالمية الشهرية لمزيج برنت بالدولار الأمريكي لخمسة سنوات افتراضية.	39
.8	**مثال5**	بيانات خاصة بعدد من موظفي إحدى الشركات تضم ثلاثة متغيرات هي مرتب الموظف والعلاوة والخصم السنوي بالدينار الليبي.	42
.9	**car_insurance_claims**	بيانات من أمثلة برنامج SPSS تحتوي على خمسة متغيرات لعينة من 128 شخصا في إحدى شركات التأمين على السيارات، والمتغيرات هي: عمر المؤمّن على السيارة بالسنوات ضمن فئات عمرية، فئة تصنيف السيارة، عمر السيارة بالسنوات، معدل قيمة التأمين المدفوعة من قِبل الشركة بالدولار، وعدد مبالغ التأمين المدفوعة.	50

.10	**أقسام كلية العلوم**	أعداد الطلبة الذكور والإناث في خمسة أقسام في كلية العلوم بجامعة بنغازي لعام 2012، والأقسام هي: الرياضيات، الإحصاء، الفيزياء، الكيمياء، والنبات.	56
.11	**المأكولات المفضلة**	آراء مجموعة من الأشخاص حول وجباتهم المفضلة، (مثلجات، لحوم حمراء، سلطات، دجاج، بيتزا)، من المطاعم المختلفة.	61
.12	**مقاومة المعدن**	نتائج مقاومة خليط المعادن قبل وبعد إضافة معدن جديد، يستخدم في بناء جدران المنشئات العسكرية، خلال فترة ثمانية أشهر.	80
.13	**معدلات الطلبة**	عينة مكونة من 90 طالبا من الطلبة القدامى في كلية العلوم بجامعة بنغازي مقسمة بحسب التخصص؛ (رياضيات، إحصاء، وفيزياء)، وتشمل معدل الطالب، (على مقياس من 0.00 إلى 4.00 درجة)، جنس الطالب، وعدد الفصول الدراسية التي أنجزها الطالب.	85
.14	**المقابلات الشخصية**	النتائج التي حصل عليها 20 شخصا أجروا مقابلات شخصية للعمل في قسم الموارد البشرية في شركتين، (التقييم من 100 درجة).	110
.15	**عينات ضغط الدم**	قياس ضغط الدم لعينة من 15 مريضا في إحدى المستشفيات قبل وبعد إعطاؤهم دواء تجريبي موسع للأوردة الدموية.	112
.16	**أسعار الشواحن**	أسعار 10 أنواع من شواحن الهواتف النقالة بالدينار الليبي في ثلاثة متاجر مختلفة.	114
.17	**مستوى الأرق**	مستوى الإنتاج ومستوى الأرق لعينة مكونة من 90 عامل في أحد مصانع الحديد والصلب.	118
.18	**فيتامين دال**	معدلات فيتامين دال (Vit. D) وأعمار عينة مكونة من 60 شخصا إضافة لتصنيفهم ذكور وإناث.	121
.19	**أطوال الأطفال**	أطوال عينة مكونة من 20 طفلا (بالسنتيمتر) تم تقسيمهم إلى خمسة مجموعات، كل مجموعة تم إعطاؤها نوع محدد من ضمن 5 أنواع من الحليب، وبداخل كل مجموعة من هذه المجموعات الخمسة تم اعتماد 4 أنواع من فيتامين دال.	141
.20	**بيانات الطلبة2**	ملف بيانات الطلبة مضافا إليه متغير جديد يمثل معدل عدد ساعات الدراسة اليومية للطالب.	148
.21	**استبيان المنتجات الغذائية**	استبيان يتضمن 10 أسئلة تم توزيعه على عينة مكونة من 100 شخص، ويتضمن استجابة هؤلاء الأشخاص لأسئلة متعلقة بجودة المنتجات الغذائية العربية والأجنبية في العموم.	221

المراجع

1. Barton, B., and Peat, J., (2014), *Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal*, John Wiley & Sons Ltd., U.K.
2. Brink, D., (2008), *Statistics*, David Brink and Ventus Publishing ApS. U.S.A.
3. Douglas, M. and George, R., (2003), *Applied Statistics and Probability for Engineers,* John Wiley and Sons, Inc. U.K.
4. Fernandes, M., (2009), *Statistics for Business and Economics*, Marcelo Fernandes and Ventus Publishing. U.S.A.
5. Field, A., (2006), *Discovering Statistics using SPSS*, SAGE Publications, London, U.K.
6. Frank, H., and Althoen, S., (1994), *Statistics: Concepts and Applications,* Cambridge University Press. U.K.
7. Gerber, S., and Finn, K., (2005), *Using SPSS for Windows: Data Analysis and Graphics*, Springer Science and Business Media, Inc. U.S.A.
8. Griffith, A., (2010), *SPSS for Dummies*, Wiley Publishing, Inc. U.S.A.
9. Gupta, V., (1999), *SPSS for Beginners,* VJBooks Inc., U.S.A.
10. Han, J. and Kamber, M., (2000), *Data Mining: Concepts and Techniques,* Morgan Kaufmann Publishers. U.S.A.
11. Landau, S., and Everitt, B., (2004), *A handbook of statistical analyses using SPSS*, Chapman & Hall/CRC Press LLC. U.S.A.
12. Leech, SN., Barrett, K., and Morgan, G., (2005), SPSS for Intermediate Statistics: Use and Interpretation, Lawrence Erlbaum Associates, Inc. U.S.A.
13. Moore, D., (2003), *The Basic Practice of Statistics,* W. H. Freeman Publishers. U.S.A.
14. Spiegel, M., Schiller, J. and Srinivasan, R., (2001), *Schaum's Easy Outline of Probability and Statistics,* The McGraw-Hill Companies, Inc. U.S.A.
15. Stephens, L., (2006), *Schaum's Outline Beginning Statistics,* The McGraw-Hill Companies, Inc. U.S.A.
16. University of Bristol, (2010), *Introduction to SPSS (version 18) for Windows*, Information Services, U.K.
17. Wackerly, D., Mendenhall, W., and Scheaffer, R., (2002), *Mathematical Statistics with Applications*. Duxbury Thomson Learning, Inc. U.S.A.

لماذا SPSS؟

يُعد برنامج SPSS، والذي يمثل اختصارا للحزمة الإحصائية للعلوم الاجتماعية، البرنامج الإحصائي المفضل للطلبة والباحثين المتخصصين في العلوم الطبية والنفسية والبيولوجية، والعلوم الاجتماعية والأدبية، وكذلك العلوم الاقتصادية والمالية، وغيرها من التخصصات، حيث أنه يحتوي على العديد من الأساليب الإحصائية والمقاييس التي تتعلق بطبيعة هذه التخصصات وطرق تعاملها مع البيانات. ومن ضمن المزايا التي يتمتع بها البرنامج، مقارنة بالبرامج أو الحزم الإحصائية الأخرى المرونة الفائقة في إدخال البيانات وتعديلها ومعالجتها، وتنوع طرق ترميز وتحويل المتغيرات، وكذلك سهولة استيراد وتصدير ملفات البيانات من وإلى البرامج الأخرى. وفي هذا الكتاب، نقوم بشرح كيفية التعامل مع البيانات بداية من إدخالها في البرنامج وإجراء التعديلات عليها، مرورا بعرض طرق استكشافها وتمييز الأنماط المختلفة ضمنها، ثم استخدام الأساليب الأكثر عمقا في تحليلها ونمذجتها، إلى توضيح كيفية تفسير النتائج الإحصائية وعرضها بصورة مبسطة.